

THE CARITAS IN VERITATE FOUNDATION WORKING PAPERS
“The City of God in the Palace of Nations”

The Humanization of Robots and the Robotization of the Human Person

*Ethical Reflections on Lethal Autonomous Weapons Systems
and Augmented Soldiers*



THE CARITAS IN VERITATE FOUNDATION WORKING PAPERS

“The City of God in the Palace of Nations”

The Humanization of Robots and the Robotization of the Human Person

*Ethical Reflections on Lethal Autonomous Weapons Systems
and Augmented Soldiers*



With a selection of documents from the Holy See's engagement on
Lethal Autonomous Weapons Systems

www.fciv.org

Edited by Alice de La Rochefoucauld and Stefano Saldi

Electronic document preparation by Margarete Hahnen

Published by FCIV

16 chemin du Vengeron, CH-1292 Chambésy

© The Caritas in Veritate Foundation

ISBN: 978-2-8399-2266-1

EDITORIAL

Rev. Fr. Antoine Abi Ghanem and Stefano Saldi, Permanent Observer Mission of the Holy See to the United Nations and Other International Organizations in Geneva	5
---	---

SECTION ONE: *The Humanization of Robots and the Robotization of the Human Person* 13

The Humanization of Robots: Lethal Autonomous Weapons Systems and Ethics

Dominique Lambert, Professor at the University of Namur and Member of the Royal Academy of Belgium	15
--	----

The Robotization of the Human Person: From the Use of Armed Robots to the Advancement of the “Augmented Soldier”

Dominique Lambert, Professor at the University of Namur and Member of the Royal Academy of Belgium	47
--	----

The Need for an Ethic of the Responsible Person aware of his/her Limitations

Dominique Lambert, Professor at the University of Namur and Member of the Royal Academy of Belgium	61
--	----

SECTION TWO: *Documents of the Holy See on Lethal Autonomous Weapons Systems* 67

Introduction to the Documents of the Holy See

Alice de La Rochefoucauld, Director of the Caritas in Veritate Foundation	69
---	----

Statement at the United Nations Informal Meeting of Experts on Lethal Autonomous Weapons Systems (CCW)

Archbishop Silvano M. Tomasi, Permanent Observer of the Holy See to the United Nations and Other International Organizations in Geneva, 13 May 2014	73
---	----

The Use of Lethal Autonomous Weapons Systems - Ethical Questions

Permanent Observer Mission of the Holy See to the United Nations and Other International Organizations in Geneva, 16 April 2015	77
---	----

Elements Supporting the Prohibition of Lethal Autonomous Weapons Systems

Working Paper Submitted by the Permanent Observer Mission of
the Holy See to the United Nations and Other International Organizations
in Geneva, 7 April 2016 89

*Statement at the United Nations Informal Meeting of Experts on Lethal
Autonomous Weapons Systems (CCW)*

Archbishop Ivan Jurkovič, Permanent Observer of the Holy See to the
United Nations and Other International Organizations in Geneva,
11 April 2016 95

EDITORIAL

REV. FR. ANTOINE ABI GHANEM
STEFANO SALDI

Permanent Observer Mission of the Holy See to the United Nations and Other International Organizations in Geneva

A stark reality of history, and of weapons in particular, is that the fascination for rapid developments in science and technology has often outpaced thoughtful developments in human responsibility, values, conscience and international humanitarian law that such progresses ought to consider.

The case of Lethal Autonomous Weapons Systems (LAWS) is no exception. It has been argued that this could very well be a “Revolution in Military Affairs” (RMA), with respective risks to fundamentally change the nature of conflicts.

For these reasons and in order to understand the range and gravity of fundamental ethical issues and implications involved, it is of the utmost importance to enhance the discussion on autonomy in its various applications and in all segments of our society. One of the objectives of this publication is to provoke a debate to develop an informed position to establish a common understanding of the subject matter.

In current literature, one encounters a large number of classifications, based on different degrees of autonomy delegated to the machines. This is a clear symptom that the concept of autonomy is far from being unequivocal, since there is a spectrum of possible interactions between the machine and a human agent. To this purpose, in the first part of this working paper, a detailed definition of LAWS is put forward, thus making it possible to move forward by further restricting and clarifying what the usual terminology describes as “appropriate” or “meaningful” human supervision.

It is very important, therefore, to specify precisely the type of machine that is being considered. In the case of LAWS, the ethical and legal problems differ to a great extent, depending on whether the machines

are controlled, supervised, autonomous or innovative. In particular, innovative LAWS raise a variety of perplexing legal and ethical issues. How can we tolerate machines that reach beyond the objectives set by a responsible human agent?

The overarching underlying question is the following: can there be truly a respectful way to use - under certain conditions and at certain times - autonomous innovative machines?

A machine is a complex set of circuits, and this material system can never become a truly morally responsible agent. Only the human person is truly responsible for his/her actions, insofar as he/she is the source and principles of those acts and is the only who can truly answer for it.

Robots and artificial intelligence systems are based on rules, including protocols for the invention of new rules. But legal and ethical decisions often require going beyond the rule in order to save the spirit of the rule itself.

Any armed intervention must be duly considered, and its legitimacy, legality and conformity must be at all times verified with its purposes. The use of autonomous armed robots has the dangerous peculiarity of being able to eliminate an action of its content, its purposes, and its roots and easily conceal or dilute the responsibilities associated with it. Ultimately, the use of LAWS is no longer a human action but, perversely, it tries to exonerate the truly responsible agent.

Our overarching ethical criticism rests fundamentally on a de-humanization and de-responsibilization of the action by the human person since it causes a loss of content in the human action and the refusal of the responsibilities associated with it.

The idea of a “moral” and “human” war waged by non-conscious, non-responsible and non-human agents is a lure that conceals desperation and a dangerous lack of confidence in the human person. Yet, this ignores the fact that for a machine, a human person, just like everything else, is only a set of numbers, is only one being among others, interchangeable, and an object of application of certain rules or protocols. Now this is precisely the “inhuman”: the delegation of powers to autonomous machines puts us on the path of negation, oblivion or contempt for the essential characteristics unique to the human persons.

Some argue that robots could wage wars instead of humans so that the lives of soldiers would be spared. Paradoxically, however, even if one were to place oneself in the utopian configuration where two nations had robotic means, a simple symmetrical war between robots could not exist. History shows that war is the search for asymmetry and the weak point (the traditional “moral” objective in war) of a nation is always human life, in particular civilians. As is the case nowadays with terrorism, belligerents would certainly attempt to undermine human life and not simply destroy machines. Furthermore, with robots becoming increasingly less expensive, their loss will only have a minor impact. Let us imagine how much easier it would be to justify, to the national public opinion and media, the loss of a robot, instead of the loss of a soldier.

Moreover, there is another intrinsic and real risk, related to the emergence of a new potential arms race, faster and much less controllable than in the case of nuclear weapons. Compared to the rapid development of computer technology or artificial intelligence, the race for progress in the field of robotic weapons will be a source of increasing instability, urging nations that are victims of the “security dilemma” to increasingly invest more in sophisticated weapons rather than allocating resources to other sectors such as health and education. The case of atomic weapons should serve as a warning and offers a precedent for reflection here. The risks posed by LAWS, if no preventive actions are taken at the international level, would be that their immediate military interest and destabilizing effects could stimulate their development. The most advanced and developed nations would be once again favored in a new arms race, further increasing the gap between poor and rich nations.

Another worrying concern being outlined in the first part of this working paper is that it would be extremely dangerous to consider an “electronic personality” for the robot (civilian or military) or to give it a legal status as a human person. The confusion between the concepts of things (objects) and persons (subjects) risks jeopardizing the very foundation of law – that is the human person.

If we wish to remain faithful to an ethics based on the respect for the characteristics that are essential to the human person, we must certainly consider the prohibition of research and development of innovative armed robots, i.e., LAWS, the autonomy of which is maximum, without human supervision with respect to the programming, the piloting, and the learning (a situation where the human agent is entirely “out of the loop”). Such machines present an intrinsic risk that, at some

point, they may deviate from the areas of evolution or the objectives prescribed by the political or military responsible authority. Such loss or dilution of responsibility is unacceptable, for it nullifies the decision of its inherent human nature.

The human person, whether a political or military decision-maker, can in no case put himself/herself in a situation where he/she would be obliged to accept the results of actions carried out by machines, actions which he/she would not have consented to in conscience. Certainly, in this regard, the chain of command in military operations also would be significantly affected by using a stand-alone weapons systems. One could also query who, in the chain of command, will be able to oppose “decisions” taken by the machine? Will the autonomous combat machine occupy a certain place in the military hierarchy? And, in this case, will some lower-ranking military personnel necessarily and obligatorily be subject to the directives given by a robot?

To limit the “inhumanity” of war, to the extent possible, it is important to preserve, at the heart of the tragedies of armed conflicts, a certain space where people can still be sensitive to the misery of the other and regard him/her as a brother/sister and where one can risk gestures of forgiveness and reconciliation. This may seem paradoxical, but it is absolutely essential. There is a glimmer of humanity in the heart of conflicts, and this surpasses the criteria of utility and success.

A reflection on military robotics today cannot ignore the issue of “augmented” soldiers, which raises several ethical questions when looking critically at the transhumanism movement. Nowadays already, with the intense use of robotics by soldiers, we could likewise foresee a vast movement of robotization of the soldiers themselves, mainly by the augmentation of the soldier, which cannot proceed - just as in the case of innovative LAWS – without posing deep ethical and legal questions.

While ways of intrusively augmenting the performance of soldiers have existed for a long time, it is the elimination of an absolute reference to human nature that once again is problematic. The content of the second part of this working paper, therefore, rings another alarm bell: we must remain vigilant to focus on the issue of the augmented soldier, a combatant who has received the means or treatment that aim at increasing his/her performance.

For instance, such enhancement could relate to endurance or physical strength as well as speed of movement, to physiological resilience or the

modification of the body, or it could be envisaged from the perceptual point of view or it could apply to cognitive ability, decision-making, positioning and communication.

The presence of nanorobots capable of controlling or intervening in physiological parameters inside the bodies of soldiers could also exemplify this type of augmentation. In certain augmentation projects, the human person is transformed into an object. The robotized or augmented human person progressively becomes an instrument, a tool for combat. Eventually, the difference between soldiers and their equipment is erased and they become equipment themselves. This creates a major ethical problem. Effectively, the soldiers lose their freedom of decision.

We need to keep in mind that the body is not simply an object; it is the source of expression of a person, and thus the reflection of who we are. Therefore, combatant augmentation must be regulated by the requirement to preserve the autonomy of judgment and action. We need to be careful that augmentation does not transform soldiers into unconscientious “cyber puppets”.

It could be argued that augmentation that would control human parameters and quickly heal soldiers could be beneficial. Nevertheless, we cannot rule out that these systems very quickly could turn against the soldiers’ health, if such devices are susceptible to being controlled from the outside by an irresponsible authority.

It ought to be strongly emphasized that not all “augmentations” are opposed, just like the case of LAWS on which a human agent retains supervision, but only those that objectify or enslave the human person. For instance, one might think that exoskeletons that allow disabled people to move, devices that help to carry heavy loads, or all types of prostheses that contribute to simplify our existence and increase our quality of life, are to be welcomed and further developed.

The suggested ethical framework regarding the issue of augmented soldiers, therefore, begins with the respect for the integrity of the human body. Any augmentation of combatants should be done in a way that is reversible and respects the human body. Objectifying and artificially enhancing the body in a way that could be dangerous is *contra naturam*.

Hence, another way to envisage augmentation of soldiers is proposed: a moral augmentation. The authentically augmented individual should

be a human person endowed with a greater awareness of his/her responsibilities. Regrettably, it is not certain that the ethical education of combatants is keeping pace with their technological augmentation. Since the progressive elimination of the human body and dignity is associated with the transhumanist and “augmentative” mentality, this probably will lead us to the loss of shared core values, on which we would have expected to base relevant international laws and at least a minimum amount of shared morality. Such concealment of the human body and disruption of human interactions could easily lead to a decline of the *conscience publique* and of moral standards.

When we face the challenges brought about by the robotization of warfare, while a legal framework is certainly needed, this will not be enough: we will also need an appropriate ethical reflection profoundly engrained in what constitutes the richness and *grandeur* of the human person.

This profound attention to the human person, this respect for his/her own limitations, considered as assets, could lead to the establishment of a framework underpinning future discussions concerning the evaluation of these new military technologies, without disregarding those advances in science and technology that could guarantee health, well-being and peace.

Some questions that have emerged from this reflection are suggested as reference points for future discussions. These questions could constitute a template for an “ethical check list” that allows us to highlight and carefully consider a series of crucial dangers that should not be ignored.

While it is clear that there could be many answers to such questions, it may be important to consider how a common ground, based on the respect of human dignity, could be sought in the ongoing discussions. There is not the slightest doubt that the way forward is indeed the path of intellectual debate at all levels: among diplomats, politicians, scientists, philosophers, lawyers, military, businesses etc... The issues raised by innovative LAWS and augmented soldiers exceed by far the scope of the conduct of war but have implications and consequences on the entire human family. Therefore, it is necessary to act and use technologies in a suitable way so that these actions remain at all times compatible with the human person, with his/her body and spirit, with his/her strengths but also his/her limitations.

In conclusion, international security and peace are best achieved through the promotion of a culture of dialogue and cooperation, not through an arms race. Thus, an “ethics of fraternity” should be built on human relationships, on exchanges where compassion, understanding and respect for values must take precedence over profit. A world in which autonomous machines are left to manage, rigidly or randomly, fundamental questions related to the lives of human beings and nations, leads us imperceptibly to dehumanization and to a weakening of the ties that underlie the possibility of a true and lasting fraternity.

SECTION ONE

THE HUMANIZATION OF ROBOTS AND THE ROBOTIZATION OF THE HUMAN PERSON

**ETHICAL REFLECTIONS ON LETHAL AUTONOMOUS
WEAPONS SYSTEMS AND AUGMENTED SOLDIERS**

THE HUMANIZATION OF ROBOTS: LETHAL AUTONOMOUS WEAPONS SYSTEMS AND ETHICS

A Caritas in Veritate Foundation Report by

DOMINIQUE LAMBERT

Professor at the University of Namur and Member of the Royal Academy of Belgium

1. A definition of the robot

A robot is a system comprised of three mutually interdependent components. First, **sensors** that allow the acquisition of information about a given environment. Then, **processors** that organize the information newly received or already available through artificial intelligence programs. Third and finally, **effectors** that allow action on the environment in which the system is immersed. It is further assumed that this system has the capacity to operate partially or totally without the mediation of a human agent. Therefore, partial or total independence from a human agent is essential to the definition of a robot.

Delving into a further classification, one can distinguish **physical** robots from **electronic** robots (bots).

Physical robots, on the one hand, are systems embedded in a physical structure that are capable of moving in one or more specific geographical environments (terrestrial, aerial or submarine). They can, in some cases, take on the appearance of human beings (android robots). In the case of physical robots, it is generally assumed that they are potentially recoverable after an action. For example, a cruise missile does not fall into this category of systems.

On the other hand, electronic robots (bots) are systems that “evolve”, largely autonomously, in cyberspace. They are able to extract information from electronic networks, to process them and then to act on them. It is important to consider such systems in the new forms of cyberwarfare. An example of such **bots** in the civilian sphere is given by financial robots that can carry out a very large number of financial transactions without the mediation of human operators.

**Partial or total independence
from a human agent is essential
to the definition of a robot.**

The weapons systems referred to below may be considered as physical or electronic robots within the meaning of the definition above-mentioned, which, however, are also capable of autonomously delivering a physical or an electronic action, be it lethal or not, for offensive or defensive operations.

2. The question of the definition of LAWS

A weapon system is said to be autonomous according to the United States Department of Defense if, once activated, it can select and engage targets without the intervention of a human agent.

It is important to put forward a clear definition of a Lethal Autonomous Weapon System (LAWS).¹ A number of publications have focused on possible definitions of autonomy. According to the United States Department of Defense², a weapon system is said to be **autonomous** if, once activated, it can select and engage targets without the intervention of a human agent. The definition could be further extended by saying that this is a programmed system for carrying out a series of important tasks, such as identifying, monitoring, prioritizing, selecting and combating targets without direct human supervision. We say “without direct human intervention” since, in this case, human agents have programmed such a system to perform precise tasks.

Indirect human supervision can be envisaged and realized in two different ways. First, even if the machine can perform certain tasks without a human agent, it is a human person who fixes the geographical limits in which it can operate and which determines the set of authorized behaviors and actions. The human person, for instance, can determine an aerial or underwater area where the system can perform. This restriction will prevent any action in an environment where there is uncertainty over the presence of non-combatants. Furthermore, the human agent can also impose the limits of a possible action by providing for an automatic interruption of the machine or its return under the direct supervision of a human agent.

Indirect supervision may also be employed by writing programs that define and constrain the set of all possible actions of the machine but not the particular actions that the machine itself will “choose” in one or another environment. One can imagine an even broader definition of autonomy that would further diminish the scope of human supervision. This would be, for example, the type of autonomy characterizing, on the one hand, **self-learning machines** or, on the other hand, **self-programmable machines**, which we may refer to as “innovative” machines.

Self-learning machines are capable, for example, of detecting correlations between data collected in a particular environment, of developing new characteristics or categories and taking them into account for future actions. Such techniques are already implemented in **data mining** and **deep learning**. Learning can happen in an environment and from data chosen by a human tutor. The latter can also – totally or partially – direct and control the learning process. However, it is also possible to envisage a learning that is totally unsupervised by a human agent.

Self-programmable machines are systems that can change their lines of code either spontaneously or in order to adapt to changing environments. “Genetic algorithms” are one of the possibilities envisaged for self-programming. A series of alterations are introduced into the initial program and an optimal adaptation criterion to a given environment or task helps to select the new program.

It is clear that “innovative” machines, i.e., self-learning or self-programmable, are systems that, at some point, may perform unforeseen actions, escaping from the predicted geographical limits and thus contradicting the objectives set by the authority responsible for their deployment.

At this point it is important to mention the difference between **automaticity** and **autonomy**.

An **automated system** is one that performs tasks without human supervision but whose actions are perfectly identified and predictable. The servomechanisms are of the following type: to a certain well-identified stimulus appeared in the environment corresponds a perfectly programmed and planned answer.

An **autonomous system** necessarily gives way to certain unpredictability of actions, which is what sets it apart from **automated systems**.

Nevertheless, we can think of systems whose entire set of behaviors is perfectly identified, without one being able to predict the behavior that will occur in a particular, unforeseen situation. This is the case of many situations studied by the **probability theory**. For instance, it is known that a coin can fall only on the tail side or on the head side (all the possibilities of behaviors are clearly known), but it is not known, during a particular throw, which side will face up. This type of situation could correspond to an example of autonomous systems indirectly supervised by a human agent. This supervision could amount to fix, in the program, all the possible behaviors of the machine, knowing the probability of each such behavior indirectly supervised (but accepting not knowing which precise behavior will occur in a specific situation).

Innovative autonomous machines, however, are those in which we have both an unpredictability of “local” behavior (those that occur in a particular situation in a given environment) and an unpredictability of the totality of their potential behavior and the limits of their field of action.

The ethical and legal problems are very different depending on whether the machines are either automatic or autonomous, non-innovative or innovative. It is, therefore, very important to specify precisely the type of machine that is being considered. In current literature, we can find a large number of classifications based on different degrees of autonomy left to the machines³. Such a variety indicates that the notion of autonomy is not unequivocal, since there is a whole spectrum of possible interactions between man and machine, from programming, to controlling and learning.

An automated system is one that performs tasks without human supervision but whose actions are perfectly identified and predictable.

An autonomous system necessarily gives way to certain unpredictability of actions, which is what sets it apart from automated systems.

Innovative autonomous machines, however, are those in which we have both an unpredictability of “local” behavior and an unpredictability of the totality of their potential behavior and the limits of their field of action.

For each interaction, we can envisage a more or less important contribution of the human or robotic component explaining the multiplicity of the classifications encountered in the literature.

For the purpose of this paper we suggest – far from wanting to be exhaustive - the following classification based on four types of processes in which the human agent could be engaged when using a machine. (1) The process of **programming**, (2) the process of **piloting** (control of behaviors, movements, actions, etc.), (3) the process of **supervision on the piloting** (the piloting being carried out by the machine itself) and (4) the process of **learning** (if such a process is possible). The human actor can, in relation to these four types of processes, either (a) maintain full and permanent control (i.e., “**human in the loop**”), or (b) just maintain a partial, indirect or intermittent control (i.e., “**human on the loop**”), or (c) to delegate, completely and permanently, the mastery to the machine (i.e. “**human out of the loop**”), except perhaps in terms of programming or learning, which would remain under human control.

Summing up what has been outlined above, we obtain the following table, where “Yes” refers to the fact that the human agent is engaged in the process and “No” that he/she is not engaged (temporarily or permanently).

TABLE 1

	Supervision on the Pro- gramming	Piloting	Supervision on the Pilot- ing	Supervi- sion on the Learning
« In the loop “	Yes	Yes	Yes	Yes
« On the loop “	Yes/No	No	Yes	Yes/No
« Out of the loop “	Yes/No	No	No	Yes/No

A totally tele-operated machine corresponds to the first row of Table 1. It is clear here that piloting (human) involves supervision (human). This is the case of most aerial drones used today.

In the second row of the Table we can first have machines whose programming is completely fixed by the human agent. But the latter is not piloting the machine. He keeps a look, “from outside” at the behavior of the machine, while retaining the ability to modify or interrupt it. However, the machine can operate by itself and “choose” this or that type

of action without direct recourse to a human agent. This is the basic case of a supervised machine with human intervention during the programming process.

But we could also think about machines that are supervised “from outside”, with respect to the piloting, and at the same time are self-programmed. The latter refers to machines without human intervention with respect to programming.

Furthermore, if the machine has also learning capacity, the human agent may or may not control it. One can imagine, in fact, a **supervised** machine, with respect to the piloting, but whose real-time control could be influenced by a process of “machine learning” escaping the human agent. The latter case would fall into a category of partially supervised robots (i.e., supervised from the point of view of piloting, but not from the point of view of learning).

In the second row of Table 1 we could also envisage machines for which the human agent would be “on the loop”, that is to say, would retain capabilities for supervising piloting, but which would possess self-programming capabilities, with or without self-learning capability. Here again we would have another example of a partially supervised machine (i.e., supervised from the point of view of piloting but not from the point of view of programming).

These cases, where the human agent oversees the behavior of a robot from outside but does not have control over its programming or its learning could be described as a case of **weak supervision**. On the contrary, when the programming and the learning of the robot is permanently supervised by a human agent, we speak of **strong supervision**. This makes it possible to further clarify what the usual terminology describes as **appropriate** or **significant** human supervision. A significant supervision has to be at least a “strong” supervision (supervision with respect to the piloting as well as indirect human intervention with respect to the programming and the learning process).

The last row of the Table above describes either machines that are escaping the human agent in terms of piloting and supervision, but which remain programmed by the human agent and subjected to it during the learning phases or machines that totally escape the human agent in terms of programming (self-programmable machine), piloting, supervision and/or learning (self-learning machine). Let us recall that when human agents are losing control of the machines’ programming and learning processes, they are qualified as “innovative”.

Hence, based on these distinctions, we may summarize by identifying four types of robots: **controlled, supervised, autonomous and innovative robots**.

A robot is “**controlled**” if the human agent has complete control over its programming, its piloting and its learning abilities.

Let us recall that when human agents are losing control of the machines’ programming and learning processes, they are qualified as “innovative”.

Hence, based on these distinctions, we may summarize by identifying four types of robots: controlled, supervised, autonomous and innovative robots.

A robot is “**supervised**” if the human agent, though not piloting the robot, still has an opportunity to intervene in the piloting, when deemed appropriate. A robot is **strongly supervised** if the human agent retains full control over the programming and the learning of the supervised robot (indirect human intervention). The robot is **weakly supervised** if the human agent does not have control over its programming or its learning (no indirect human intervention).

A robot is “**autonomous**” if the human person has absolutely no mastery of the piloting, even by outside supervision (but possibly retains control over the domain in which the robot operates and the *a priori*, deterministic or statistical knowledge of all its possible behaviors). The robot is **weakly autonomous** if the human agent retains control over its programming and its learning (indirect human intervention case). The robot is **strongly autonomous** if the human agent does not have the mastery of its programming or its learning (no indirect human intervention).

A robot is “**innovative**” if there is no human supervision on the programming or on the learning process. In this case, human agents do not have the deterministic or statistical control of all the possible robot behaviors. Let us note that a weakly supervised or strongly autonomous machine is necessarily innovative since human agents lose control at the level of programming or learning which implies an immediate loss of knowledge of all possible behaviors of the robot. Innovative robots can be either machines that are nevertheless (weakly) supervised or machines that are (strongly) autonomous. But we have to take into account in legal and ethical questions the cases concerning robots that could be (weakly) autonomous without being innovative devices.

Robots obviously have information storage capabilities. For robot applications in the fields of medicine, personal assistance or safety, for example, it is important to specify whether or not the machines will be able to store and transmit information of a private or confidential nature, drawn from the environments in which they operate. Different ethical or legal issues will arise depending on whether or not the robots have this capacity and authorization to store this private information. In the case of innovative machines, the issues concerning the information storage and diffusion are crucial.

Before moving forward, a brief epistemological remark is needed. The difference between an innovative machine and a human person lies, among other things, in “creativity”. To “innovate” (in a robot sense) and to “create” (in a human sense) are not the same thing. Indeed, a machine can generate new information; it can produce new objects and new production processes. However, it always does this by following rules and programs, including programs of modification of programs. The human person is “creative” insofar as he/she always has a sort of advance on the machine, a capacity to transgress or to invent other rules giving rise to totally new situations.

The difference between an innovative machine and a human person lies, among other things, in “creativity”. To “innovate” (in a robot sense) and to “create” (in a human sense) are not the same thing.

It is therefore important to measure the distance between an autonomous machine equipped with sophisticated programs of artificial intelligence and a human being who is certainly less rapid, that actually has less memory than a robot, but that always has, we could argue, this unique ability to step back and radically change all existing frameworks or points of views.

This point had already been highlighted in 1950 by Father Dominique Dubarle, who, in a discussion with Norbert Wiener, glimpsed in a visionary way to the possible progress of robotics. On the irreducible difference between the machine and man, Father Dominique Dubarle said⁴:

“[...] in spite of their powers, thinking machines are not and cannot be equivalent to the mind thinking itself. [...] In order to prove this, it is often enough to say that the difference between the machine and the mind is that the machine, before it exists, must necessarily be conceived by the spirit which fixes its powers and assigns to it its rules of operation. The machine is an extension, if necessary, of the human instruments of thought. But it is only this prolongation and it “thinks” only by virtue of a human thought outside of which its function loses all true significance, in order to be nothing more than a bizarre play of material energy within a complicated device”.

However, for Father Dominique Dubarle, this argument of “temporal” antecedence of human thought on the machine must be complemented by the following property of human thought which is⁵:

“[...] the emergence of more general frameworks of possibilities in relation to any particular data. It broadens the possible by its own act. [...] It is what gives the spirit its inventive character and its symbolic power, whereas the machine does not attain true invention and holds symbolic power only through delegation”.

The human mind is therefore the power that can always go far beyond the particular and the programmed to risk a real invention. There are, of course, internal limits of formal languages and of the “mechanization” (algorithmization) of thoughts used to program machines. But some of these limits could be circumvented by machines of a new kind. Therefore, the fundamental difference between man and machine is “the energy that gives birth to the possible and the creative energy that invents the decided choices of being”⁶. The innovative character of the machines of which we have spoken is not this creativity, for the machine can innovate only by prescribed rules of innovation. It is important to note also that the imitation by a robot of human capacities never means an identity between the man and the machine: to imitate is not to identify. Today, “Turing tests”, which try to point out a difference between man and machine, have developed to take account of more facets of human intelligence. However, if a machine succeeds in passing all these new tests, we could only deduce

It is important to note also that the imitation by a robot of human capacities never means an identity between the man and the machine: to imitate is not to identify.

that the machine is an imitation of the human being and not a copy of the human individual itself.⁷ The same goes for robots who try to imitate human emotions. It is interesting to note that some major social robotics specialists, such as Laurence Devillers, recommend that we do not lose sight of and eliminate this difference between the robot and the human being, while seeking to push very far the positive aspects of the interaction between man and machine⁸. It is in this line that we intend to situate our reflection.

It is important not only to reflect on the definition of the concept of autonomy and on the whole spectrum of its meanings, but also on its unavoidable and sometimes beneficial character. In some areas, it can be imagined that the existence of robot autonomy is a guarantee for the safety of people and goods. For example, in the event of severe physical or mental breakdown of pilots of a military or civil plane, it may be necessary for a machine to detect dangerous behaviors and, autonomously, resume piloting of the aircraft to avoid a crash. In the case of space research, a series of autonomous robots will certainly provide interesting solutions for exploring areas that are inaccessible to human beings. Their use will be necessary to save human life and health.

These examples show that we must not be **technophobic**. On the contrary, the study of machines capable of operating autonomously while serving and respecting the human person is to be encouraged. As will be gradually seen in this paper, it is **what we make of the autonomous mode and its ends** that raise legal and ethical challenges rather than autonomy as such. The essential question is: can there be a truly respectful way for men to use, under certain conditions and at certain times, autonomous, innovative machines?

3. The problems raised by the LAWS

A. The risk of unpredictability of innovative LAWS

Innovative LAWS in the sense defined above, raise a variety of perplexing legal and ethical issues. How can we tolerate machines that reach beyond the objectives set by a responsible authority? As highlighted in the discussions in the framework of the United Nations Convention on Certain Conventional Weapons (CCW), “a meaningful human control” in relation to these weapons systems is essential⁹; in this sense “meaningful” refers to “effective”. In fact, no responsible State would accept those kinds of weapons systems. These should therefore be banned. It would also be at odds with the *jus in bello*, since unpredictability could be translated into behavior targeting civilians to maximize military interest, in direct opposition to the **principle of distinction**.

In some areas, it can be imagined that the existence of robot autonomy is a guarantee for the safety of people and goods.

The study of machines capable of operating autonomously while serving and respecting the human person is to be encouraged.

At this stage, it is important to respond to an objection. Some would like to reject a *moratorium* on weapons, even if they are autonomous, because of the positive effects that related research might have for humanity. However, it must be clear that promoting research on the innovative capacity of machines in the civilian field does not contradict a strict ban on the development of such military-based systems. It is incorrect to say that we must not muzzle the research on strongly autonomous LAWS so as to allow its development in civil applications. And, if it is argued that these systems could be as important and interesting in the military field, it must be argued that no military leader or political leader could reasonably take the risk of using intrinsically unpredictable weapons.

We should still prevent an objection. It could be argued that the innovative robot wants to get as close as possible to human behavior which is itself innovative and creative. However, this means forgetting that the creativity of the human person is quite different as we have seen above. For example, the human being can transgress some of his/her own rules. But in order to do this, he/she must have perceived the meaning of these rules and their meaning, which requires semantic capacities that the program of a machine does not have.

Finally, it is important to note that the problem of unpredictability, which we have pinned down here, must be extended to any system capable of injuring a person. In the same way, the use by innovative robots of means, such as tasers, tear gas diffusers or incapacitating gas diffusers, should be avoided in military policing operations.

B. The risk of deviation of self-learning machines

If machines are given learning abilities without specifying the environments in which they must “form” and “inform” themselves, in other words, if one does not guide their learning, one risks, as in any other process of education or indoctrination, to see the machine “learn” and to consider as “normal” (in the statistical sense of the term) behaviors that the law or morality would prohibit. To develop moral behaviors, it is not enough to observe the “average” behaviors, in the probabilistic sense, of the population in which one is immersed.

It has been shown that self-learning machines confronted with recurring discourses that are unworthy of humanity (racist, bellicose, etc.), have included this kind of discourse¹⁰ fairly quickly as usual behavior. We must therefore be very careful from this point of view, since bots and electronic robots already exist today, and they are able to write, without human mediation, small articles for magazines which, in certain cases, have a decisive influence on public opinion.

However, it must be clear that promoting research on the innovative capacity of machines in the civilian field does not contradict a strict ban on the development of such military-based systems.

C. The risk of loss of power through fascination

The performance of innovative autonomous robots (speed of action, information processing, etc.) can dazzle its users and lead them to abandon the ability of human decision-making to the benefit of the machine. This could irresistibly lead to emptying human action from its content and its own ends. The financial sector pushed us to reflect on it. High-frequency financial transactions (e-trading) rob the real economy of its substance in favor of a purely virtual economy. Several governments have already taken action in this respect, by for example regulating the speed of transactions.

The loss of power and the fading of the content of actions lead to a sort of paradoxical situation in which the human agent develops machines to help him/her while becoming progressively enslaved.

The question arises whether one can accept that a **machine imposes** (and not only advises), at a certain moment, the behavior to a political or military authority? The answer is negative, for the machine must remain at the service of humanity and society; it can in no way become a kind of decision-making authority that eclipses the true meaning of human action. The human persons, whether a political or military decision-maker, can in no case put himself in a situation where he/she would be obliged to accept the results of actions carried out by machines, actions which he/she would not have consented to in conscience. One can also ask who in the chain of command will be able to oppose “decisions” taken by the machine? Will the autonomous combat machine occupy a certain place in the military hierarchy? And, in this case, will some low-ranking military personnel necessarily and obligatorily be subject to the directives given by a robot? The chain of command will certainly be significantly affected by the use of stand-alone weapons systems. However, a classic chain of command is essential to never lose sight of the aims of the actions and to identify those responsible.

The fascination with machine performance can have another consequence: it is the progressive forgetting of the hypotheses that structure the construction of algorithms. In the aftermath, it is surprising that a machine may have led us into an aberrant or morally delicate situation, but we often forget the presuppositions, the assumptions that govern the writing of the programs. This writing takes into account a series of preconceived ideas and particular intentions and do not take into account a series of situations that do not enter into the representations of the designers. Robotics or civil artificial intelligence are often confronted with these “omissions” linked to **technological fascination**. A classic case is given by risk assessment programs that “forget” or neglect to take into account events that are statistically very rare but whose consequences are catastrophic¹¹.

For the machine must remain at the service of humanity and society; it can in no way become a kind of decision-making authority that eclipses the true meaning of human action.

A classic chain of command is essential to never lose sight of the aims of the actions and to identify those responsible.

D. The fear of a high-intensity saturating attack

One might be fascinated by the potential capabilities of LAWS to cope with large-scale, high-intensity robotic or other attacks, saturating for instance all of a State's defenses with swarms of weapons systems. Indeed, human agents would be unable to respond adequately to these types of attacks arriving simultaneously on all fronts. A State might have the temptation to store armed autonomous robots in order to dissuade its opponents from such an attack. The use of LAWS as a force multiplier must be carefully studied. Indeed, as in the case of nuclear deterrence, nations risk entering into a sophisticated arms race, which in the long run risks introducing dangerous instability at the global level. One might want to justify the development of stand-alone weapons for these reasons of deterrence, without realizing that in the end this could create potentially "explosive" situations especially if the weapons fall into the hands of irresponsible non-State actors.

It seems relevant to seriously consider this situation. Indeed, it is fairly straightforward to arrive at a consensus that we would not want weapons that would turn against us or do not perform what is expected by a responsible authority. Totally autonomous machines, innovative ones, considered for precise tactical uses, should also be excluded for the strictly operational risks they entail. As such, it will be much more complicated to reach a consensus banning weapons presented as a means of confronting or deterring nations from carrying out very violent attacks using a multitude of weapons or swarms of robots (maybe also autonomous innovative or not). What must be considered here is the danger of an arms race with the instability it could create. The case of atomic weapons should serve as a basis for reflection in this field. In this context of the deterrence of **high intensity saturating attacks**, the risks posed by LAWS, if no preventive actions are taken at the international level, would be that their immediate military interest could stimulate their development, in defiance of the fact that it would have been shown that they would violate certain important points of International Humanitarian Law.

Indeed, as in the case of nuclear deterrence, nations risk entering into a sophisticated arms race, which in the long run risks introducing dangerous instability at the global level.

E. The risk of concealing the true responsible

Contemporary warfare has deployed techniques to reach the enemy without being hit, such as long-range missiles or high-altitude bombing. The autonomous armed robot further extends this concept by allowing operations distant from their starting bases and separated from any contact with a human decision-maker. The autonomous robot creates a true technological screen that removes the cause and effect of an act of war. Hence, this can lead to two types of risks.

On the one hand, in case of collateral damage, it will be easy to incriminate a malfunction of the machine and to try to diminish the accountability of

the responsible authority. As the technology is complex, we can always ask for expert reports of any kind and drag-on trials in such a way that the true responsible are never convicted. It can be argued that a crime committed by an intermediary, whether human or material, does not alter the legal or ethical responsibility of the person who sponsored it. In principle, this is quite true. However, in practice, it must be taken into account that the existence of technological mediations opens up a possibility for relativization or dilution of responsibilities. And that is the risk that must be highlighted here. This problem strengthens and redoubles the difficulty of identifying those responsible, as highlighted by Paul Ricoeur, when we consider all the potentially harmful effects of our actions¹²:

“Everything happens as if the responsibility, extending its radius, diluted its effects, to the point to make the author or the authors of the feared harmful effects impossible to apprehend.”

For our purpose, we could paraphrase Paul Ricoeur by saying that everything happens as if the multiplication of robotic technological mediations obscures and dilutes the ties that unite the harmful effects of military actions and their causes.

It cannot be ruled out that there may be strategies to use armed robots that are largely autonomous and innovative to avoid direct responsibility for collateral damage. At the international level it is grave to be blamed for deliberately ignoring the presence of non-combatants in the vicinity of a high-interest military object targeted by missile fire or bombing. If an autonomous or an innovative machine has carried out these actions and caused civilian victims, it may be tempting to invoke unforeseen technological dysfunctions, hiding behind the mediation offered by complex mechatronic systems.¹³

The identification of responsibilities is made even more difficult by the connection between the robots and a whole network of objects possibly containing more or less autonomous machines, more or less tele-operated, making it almost impossible to identify “responsible individuals”. Gilles Dowek has clearly highlighted this problem, showing that it will even remain a problem to those wishing to confer rights and responsibilities on non-human agents (which we refuse)¹⁴:

“This difficulty of identifying individuals in a continuum of communicating objects is a definite obstacle to the extension of the principles of law to informatics objects such as algorithms, robots or computers”.

On the other hand, unscrupulous or desperate groups may knowingly use autonomous robots, or complex networks of connected robots, to perpetrate wrongdoing that they would not have dared to attempt directly.

Some scholars have pointed out that accepting the killing of human beings by means of an autonomous machine may be unworthy of the

The identification of responsibilities is made even more difficult by the connection between the robots and a whole network of objects possibly containing more or less autonomous machines, more or less tele-operated, making it almost impossible to identify “responsible individuals”.

human beings¹⁵. But this is often disputed by the fact that it is equally unworthy for the human person to carry out high-altitude bombing or attacks by means of chemical weapons or nuclear weapons at very long range. It seems that the underlying problem is linked to a growing distance between the one responsible for the attack and his victims. This can create a kind of feeling of impunity or unconsciousness. In such a case, an attack is provoked without actually seeing the consequences, without being affected by material damage, suffering and death. With autonomous robotic weapons this problem is redoubled, because not only is a distance established, but the link, the “physical” contact, is voluntarily cut off between the initiator of the combat and the weapon he implements. The severing of physical contact in no way removes responsibility, but, like geographical distance, gives an impression of causal distance which is of great gravity. In the use of LAWS, there is something that could be hypocritical: one actually wants to cause lethal effects or damage, but without giving the impression of being personally engaged in the process that causes them. The shrinking of the direct cause-effect together with an intention to harm (which is already present to a certain extent in the use of antipersonnel mines) are one of the most problematic elements of autonomous robotic weapons.

F. The risk of an algorithmic pseudo-ethics integrated into the LAWS software

A number of robotics theorists have proposed to introduce LAWS software programs that they believe can satisfy legislation, rules of engagement, and ethical principles and standards¹⁶. A whole study has been carried out to show that the application of laws or principles requires an understanding of the contexts and situations that entail going well beyond the potentialities of the algorithms. To characterize a fact or to apply a general law to a particular case, demands, on the part of a judge, something more than simple logic, more than the pure manipulation of formal and codified rules of reasoning. The philosophy of law has long been aware that the exercise of legal activity cannot be completely modeled by a formal logical procedure¹⁷, even if it is based on a modal logic adapted to legal reasoning such as a deontic logic¹⁸. In the case of dilemmas, which are the most problematic situations for ethical judgment, a machine generally does no better than a human person. Both have to face a situation where there is no automatic rule of decision or evaluation.

It often happens that the construction of ethical algorithms requires a prior choice of the principles that will underlie their conception. Nor is it evident that all ethics can be implemented in an algorithmic way. Those ethics are the only ones whose fundamental concepts could be “quantified”. However, this immediately raises questions.

Let us suppose, for example, that we adopt a **consequentialist ethic**. It is not that easy to assess the consequences of a given act and to know

The shrinking of the direct cause-effect together with an intention to harm are one of the most problematic elements of autonomous robotic weapons.

where to stop in the future in order to decide that we took into account the future effects of a present act. In addition, it is not easy to decide on the basis of what level of risk a particular behavior or action is prohibited. Risk levels can be calculated according to the assumptions underlying the mathematical models used. However, the decision to accept or refuse this or that level of risk is something that cannot be translated as such into algorithms.

The only ethic that can really be implemented seems to be **utilitarianism**, since it is based on a calculation of maximizing “well-being” and minimizing “ill-being”¹⁹. Apart from the fact that these notions are arbitrary and relative to particular groups or individuals and that they involve a quantification that is not at all evident (especially when it comes to the lives of human beings), calculations of optimization that the algorithmic ethic implies, risk to confront the limits encountered in any **complex optimization problem** (which does not always have an algorithmic solution). The internal limits of formal systems can prevent – even in ideal situations – from finding the ethically optimal solution by calculation or by algorithm²⁰.

Instead of starting from *a priori* principles, one could, more pragmatically, rely on rules induced by a learning algorithm (machine learning algorithm). But the problem is whether to know in which environments and for how long the machine should perform its “learning”?²¹ Another problem is whether learning will be supervised by a human authority or not, and if so, what are the ethical references of the teacher? The latter could adopt the fundamental principles of bioethics: do good (**principle of beneficence**), do no harm (**principle of non-maleficence**) and be fair (**principle of justice**). However, it should need to establish a hierarchy between these principles and a quantification of their satisfaction. These needs may appear arbitrary or impossible.

In addition, assuming that the machine has been able to discover, by induction, “principles” satisfactory to our eyes, one might find oneself in situations of dilemma where one could not necessarily know which principle to use. Somewhere the necessity of recourse to the human person who chooses and who decides shows his/her irreducible necessity.

The risk of legal and ethical programs is to suggest that if LAWS contain such “ethical” or “legal” programs, they will become authentic “**moral machines**”²² that can replace a conscious and responsible human decision-maker. The risk here is that these software programs may exempt those who implemented them from their legal and ethical responsibilities. This problem is already present in civil robotics where there is a growing awareness that robots are given legal personality; a fictitious concept which allows the robot to be associated with a compensation fund for victims in the event of damage to persons or property. This is also a way of removing or even obscuring those with real accountability. This will be elaborated later on.

A machine is only a complex set of circuits and this material system cannot in any case become a truly morally responsible agent. The human person alone is truly responsible for his/her actions, insofar as he/she is the source and principle of those acts and is the only one who can truly answer for it.²³

“Moral” software is a fiction that conceals true decision-makers and their particular ethical principles, and mimics moral action that cannot be restricted to a mere automated application of rules. It is necessary to recall here that “to mimic” (to model) a moral behavior is not identical to becoming a moral agent.

In fact, robots and artificial intelligence systems are based on rules, including protocols for the invention of new rules. But legal and ethical decisions often require going outside the set of rules in order to save the spirit of the rules. Furthermore, rules and laws systems are sometimes not sufficient to know and to decide exactly what we have to do in complex and contingent situations. Traditionally, it is said that in these areas the judgment of prudence is important, which, as Aristotle had already shown, is not reduced to a logical derivation based on universal principles²⁴. This fact can be illustrated by citing Mireille Delmas-Marty’s reflection, in the context of international law, considering the place of the human person (the “thinking reed” evoked by Pascal) versus the robot²⁵:

“The assistance of robots will be indispensable to justice to master the increasing complexity of legal systems and contribute to the emergence of justice at the national, regional and global levels. But the fragility and flexibility of the reed remain all the more necessary because doubt is the condition of a justice which sometimes agrees to renounce punishing precisely “for the benefit of the doubt” or even to give up judging to allow the forgiveness that is the condition for reconciliation.”

G. Dehumanization or humanization of conflicts by robots?

The observation of Professor Delmas-Marty introduces an aspect that is often discussed in the ethics of armed robotics. It points to the irreplaceable character of the human person in the process of legal decision. In the same way, this irreducible place of the human person could be maintained in the framework of moral judgment. But some will object to the following idea. Would it not be better to fight with machines that act “coldly” without the spirit of vengeance, without hatred, rather than with soldiers whose anguish, fear or anger can lead to ignoble acts of barbarism and destruction?

Some engineers²⁶ argued that a properly programmed machine would not violate or retaliate against civilians. This may be true in some cases, although armed machines may also malfunction and develop “immoral” behaviors due to excessive consideration of **military utility parameters**.

A machine is only a complex set of circuits and this material system cannot in any case become a truly morally responsible agent. The human person alone is truly responsible for his/her actions, insofar as he/she is the source and principle of those acts and is the only one who can truly answer for it.

Would it not be better to fight with machines that act “coldly” without the spirit of vengeance, without hatred, rather than with soldiers whose anguish, fear or anger can lead to ignoble acts of barbarism and destruction?

But what is more serious here is a kind of loss of confidence in the human person.²⁷ We no longer believe in the ability of human beings to behave in a respectful way. A moral and legal education of leaders and soldiers could prove very effective, even in critical situations, to avoid all kinds of unethical behavior (torture, rape and other degrading treatment). This moral education can also make a glimmer of humanity shine in the heart of conflicts, surpassing the criteria of utility and success. If, in the spirit of what Paul Ricoeur has taught us, we must recognize that the magistrate is the one who humanizes law, the well-trained soldier, can also be the one who keeps the human person at the heart of this inhuman action, which is war. It may be important to maintain confidence in a morally trained human person who, in many local situations of wars or conflicts, will serve as a “fuse” to cut off the current of blind violence subjected only to imperatives of force or utility.

The idea of a “moral” and “human” war waged by non-conscious, non-responsible and non-human agents is perhaps a lure that conceals a desperation, a real lack of confidence in the moral or an implicit denial of any reference to an ethic that goes beyond the simple optimization of the utility criteria.

The idea of a “moral” and “human” war waged by non-conscious, non-responsible and non-human agents is perhaps a lure that conceals a desperation, a real **lack of confidence in the moral** or an implicit denial of any reference to an ethic that goes beyond the simple **optimization of the utility criteria**. If one wants to “limit the inhuman” and build a “community of values” in the words of Professor Mireille Delmas-Marty²⁸, one cannot agree to delegate crucial decisions to machines. For a machine, a human person is only a number, is only one being among others, interchangeable, and an object of application of certain rules or protocols. In fact, for the robot, all being, including human, is reduced to a set of numbers. Now this is the inhuman: to enter into this reduction, in this diminution of the human person which makes him/her lose his/her unique richness and his creative capacity to transcend any predefined category. Insensibly, the delegation of powers to autonomous machines puts us on the path of negation, oblivion or contempt of the essential characteristics of the human person.

H. The lure of a war conducted with LAWS with lesser human risk

One of the risks of LAWS lies in the fact that they could easily be presented as precision weapons capable of carrying out wars at very low human cost. This is already the case for remotely operated aerial drones. However, this kind of justification does not hold ground. Indeed, on the side of the country attacked, there will necessarily be victims and among them there may be civilians. The absence of victims on the side of the attackers will very likely provoke a desire for revenge that will increase the acts of terrorism or guerrilla warfare, as seen nowadays²⁹. In the attacked country, being the victim of robotic attacks, there will likely be a multiplication of strategies to conceal military installations in the vicinity of sensitive areas such as hospitals or close to civilian-inhabited areas, to serve as **human shields**. Again, the absence of risks on one side

will result in an increase in the risks on the other and especially on the civilians' side. Asymmetry in the number of victims is a factor that will necessarily fuel new violence. Even if one were to place oneself in the utopian configuration where two nations had robotic means, a simple symmetrical war between robots could not exist. History shows that war is the **search for asymmetry** and the weak point of a nation is always human life. As it is the case nowadays with terrorism, belligerents would certainly attempt to undermine human life and not simply destroy machines.

In both cases: either in the total asymmetry (present situation) or in the almost perfect symmetry (utopian situation), the unstable state of conflicts would almost always evolve towards an increase in civilian casualties (as the most easily attainable and vulnerable), which is in contradiction to the spirit of International Humanitarian Law³⁰.

I. Side effects of fully autonomous weapons: fear and induction of harmful behavior

The use of **complete autonomous innovative machines** can, moreover, create a feeling of arbitrariness. The attacker who knows that his aggressor uses fully autonomous machines may have the impression that his fellow citizens are likely to be the target of random, completely arbitrary attacks. This can create a sense of fear among the civilian population. This factor must be taken into account in the use of these weapons. A nation that decides to use weapons specifically designed to terrorize non-combatants (hoping to have indirect effects on combatants and their leaders) would of course contradict the foundation of International Humanitarian Law.

These "side effects", which are important, have not been adequately studied³¹. Even if a weapon does not directly deliver its lethal force it may have a harmful effect on both sides of the belligerents. Let us consider the situation with the use of drones (autonomous or not). People can feel a threat above them and also a sense of being under permanent surveillance. We are already experiencing intrusion or non-privacy when we learn that by using our credit card or GPS, systems with **deep learning capabilities** that can track us and, by crossing huge masses of data, learn a lot about our private life. Imagine the feeling of knowing that in the sky above us there could be autonomous systems capable of following each of our movements and entering intrusively into our lives. We must not minimize the fears that may exist in people who do not dare to travel or to gather, out of fear of espionage or intervention whose risk seems omnipresent. This psychological burden is likely to disrupt a return to peace after hostilities or to provoke feelings of vengeance that may lead to bellicose actions.

This fact must be taken into consideration in a *jus post bellum* thought. Indeed, the technology of armed robots or surveillance can plunge nations into an atmosphere of continual mistrust. Robots, by their furtive and

History shows that war is the search for asymmetry and the weak point of a nation is always human life.

Today, the technological possibilities of robotics could tempt the powers to continue to mark their hold by a kind of remote and ongoing occupation, a new form of occupancy of physical or electronic space; thus inducing an absence of a real withdrawal of forces necessary for the restoration of confidence and peace.

It is therefore necessary to reflect on the fact that the technological possibilities and the facilities or temptations which they present for the execution of certain harmful actions have an important ethical impact.

intrusive character, can carry on a sort of ambiguous but effective occupation of territories. Formerly, the withdrawal of troops of occupation meant a kind of return to a mutual confidence. Today, the technological possibilities of robotics could tempt the powers to continue to mark their hold by a kind of remote and ongoing occupation, a new form of occupancy of physical or electronic space; thus inducing an absence of a real withdrawal of forces necessary for the restoration of confidence and peace.

This brings us to a point that seems essential in reviewing the ethics of LAWS. It is often said that ethical problems arise only in the case of the use of technological tools, but that these remain neutral from an ethical point of view. It is true that a knife can be used for the good or to commit a misdeed. The knife, because of the risks it can present, requires a minimum of precautions such as the fact of not leaving it within the reach of children or irresponsible persons for example. Not taking such precautions or not paying attention to them would be a kind of foresight that, ethically, already poses a problem. It is therefore necessary to reflect on the fact that the technological possibilities and the facilities or temptations which they present for the execution of certain harmful actions have an important ethical impact. You may not use miniaturized drones to make incursions into foreign territory. But the ease with which one can introduce oneself into this territory without being noticed too much can constitute an irresistible temptation for the user, especially during critical periods of tension. One must think of technology as a potential inducer of certain harmful behaviors or temptations.

Just as the fascination for the performance of a machine can dazzle a human agent to the point of seeing him irresistibly delegate his most important powers without any critical thinking, likewise, the potential performance and facilities offered by machines can induce an irresistible temptation in human beings to use them. If it is for the good, we do not have to worry about it, but in the case of certain robotic weapons, an ease of incursion and discreet operation, without risk of loss of life, could very well push a nation to use them.

Today we must reflect on the ethical consequences of the ease of implementation of certain potentially very dangerous military technologies. The conjunction of this ease and these potentials almost inevitably leads to their use, especially in geopolitically critical situations. However, to induce a temptation with the full consciousness of what one is doing is ethically wrong.

Another aspect of this phenomenon of inducing harmful behavior could be to make a nation believe that it is invincible or superior because of the possession of sophisticated weapons and not involving the direct loss of its soldiers. There is propaganda or **luring effect** induced by certain military technologies which can thus be very harmful.

J. The risk of proliferation of armed robots and a new arms race

Information on robot technology is readily available and the implementation of this technology is within the reach of many engineering teams. It is therefore likely to proliferate and to evolve rapidly. This risks the emergence of a new form of an arms race, faster and much less controllable than in the case of conventional or even nuclear weapons³². Compared to the rapid development of computer technology or artificial intelligence, one might think that the race for progress in the field of robotic weapons will not easily stabilize, inviting nations to invest more in increasingly sophisticated weapons. This constitutive instability can be considered as one of the characteristics of LAWS technologies. The richest countries will once again be favored in this “race”, strengthening the divisions between nations³³.

The possibly small size, or use in swarm, of these armed robots will also make the task of surveillance and defense against their use by malicious groups very difficult. Therefore, a serious reflection is needed before initiating research programs on these kinds of systems. Like complex computer systems, robots are likely to be pirated and diverted, and then used outside of any legal framework. If robots are endowed with wide autonomy, irresponsible groups might be tempted to use them without the reserves and caution of nations that respect law and moral principles. The question then arises if one can really take the risk of the proliferation of autonomous weapons that could be easily diverted and used without restraint by dubious people. The proliferation and ease of implementation, as well as the potential for accelerated evolution of autonomous or quasi-autonomous armed robots, pose entirely new questions that were not raised when considering conventional weapons or at least not in the same way.

Compared to the rapid development of computer technology or artificial intelligence, one might think that the race for progress in the field of robotic weapons will not easily stabilize, inviting nations to invest more in increasingly sophisticated weapons.

4. A fundamental ethical framework

Discussions about LAWS are often limited to purely legal discussions concerning the satisfaction of the principles of International Humanitarian Law³⁴ (IHL). Even if these principles, often accepted by a large majority of States in the world, constitute a bulwark to regulate and limit violence, they are by no means sufficient. Indeed, what we should first and foremost aim at is peace between nations and this peace requires much more than what IHL implies.

True peace demands a deep reconciliation in the heart³⁵. It is true that in IHL there is reference to a certain conception of ethics. This is often mentioned, but without fundamental ethical frameworks being really brought to light. Proponents of the use of LAWS attempt to show that the arguments for opposing this type of weapon would only be very vague moral principles and without practical content.

What we should first and foremost aim at is peace between nations and this peace requires much more than what International Humanitarian Law implies.

We would like to respond to this objection by outlining some elements that could provide a genuine content for fundamental ethical reflection to assess the use of LAWS. We will base ourselves on a philosophical anthropology that describes the human person as revealing itself in his/her moral acts, as a free and responsible subject, as a social being of fraternal relationship, and ultimately as a being of compassion. This list is not exhaustive, but it aims to sketch a characterization of the human person which can help mark an appreciation of the use of technologies, especially armed ones.

From our perspective, any technology must be compatible and consistent with our conception of the human person. It would be contradictory to want to develop a technology which, on the one hand, would explicitly want to serve the human person and his/her projects and which, on the other hand and implicitly, would destroy or diminish the human person in some of its dimensions of body, mind, relations, etc. It is this **principle of anthropological non-contradiction** that will serve as a guide for an appreciation of the technologies offered by autonomous armed robotics.³⁶

A. The human person reveals itself in his/her actions

A human person is revealed in his/her conscious and voluntary acts, of which he/she bears all the responsibility.

But it is certain that his/her existence becomes quite absurd if he/she can no longer say why he/she is acting or whether he/she acts simply by responding to the injunctions of robotic systems.

A human person is revealed in (though not reduced to) his/her conscious and voluntary acts³⁷, of which he/she bears all the responsibility. The human person manifests himself/herself through his/her conscious and responsible decisions. This responsibility is specific to the human person. Hans Jonas says in this sense that “man is the only being (...) who can take responsibility for his actions, and it is precisely this power that makes him responsible.”³⁸ It is clear that the human person can and must, in certain situations, use tools, machines, and technological intermediaries. But it is certain that his/her existence becomes quite absurd if he/she can no longer say why he/she is acting or whether he/she acts simply by responding to the injunctions of robotic systems. The human existence reveals something of its meaning in its conscious and voluntary investment and actions, possibly mediated or supported by technological systems. Any individual or societal project (economic, political or social) only makes sense through a reference to human persons (and not machines). The beauty of a work, the value of an action, are ultimately measured in the bond to a human person.

Using a fully autonomous robot - in the sense of an innovative system - means emptying the action of its anthropological content and cutting off the strong link between an action and an act or decision that reveals the intentions and the heart of a person. The use of innovative robots is tantamount to suggesting that there is a human action (a duly decided military intervention) while the link to the act and to the decision is no longer present.

B. The human person is a responsible subject

The same would apply to the strategy which deliberately aims to cover or conceal the true responsibility of a decision-maker behind technological screens or by the extreme remoteness of a human agent at his service. Here, there is indeed a decided human action, but it seeks to conceal and to disguise itself in order to bear no responsibility for it. Here again, the human content vanishes not because it is absent, but because it hides itself.

One of the consequences of our ethical approach is the fact that it is extremely dangerous to consider an “electronic personality” for the robot (civilian or military) or to give it a legal status as a “legal person”. Indeed, it is a way of hiding, under a legal fiction, the true link to decision and responsibility that can only be assumed by free and responsible human persons. In the end, only the human person, free and conscious, can take responsibility for an act. As Immanuel Kant noted³⁹: “A person is a subject whose actions can be imputed to him [...] a thing is that to which nothing can be imputed. Any object of free choice which itself lacks freedom I therefore called a thing”. The confusion between the concepts of “thing” and “person” is risky from the point of view of the foundation of law. The robot is a thing and one cannot impute any responsibility to it without jeopardizing the law.

From this point of view, the COMECE Secretariat (Commission of the Bishops’ Conferences of the European Community) eloquently outlines its concern at the recommendation of the European Parliament to study the possibility of⁴⁰:

“[...] creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently.”

The COMECE’s concern is that⁴¹:

“The human person is the foundation of every legal order. For a natural person, legal personality derives from his/her existence as a human person. That personality implies rights and duties that are exercised within the frame of human dignity. Placing robots on the same level as human persons would be at odds with Article 6 of the Universal Declaration of Human Rights, which states that “Everyone has the right to recognition everywhere as a person before the law”. The Parliament’s proposal also contradicts the very concept of responsibility, based on ultimate human rights and duties. Responsibility rooted in legal personality shall only be

The disappearance or concealment of the human actor is serious from the point of view of the foundation of law and the ethics of responsibility.

exercised in presence of a certain capacity for freedom. Freedom is more than autonomy.”

The disappearance or concealment of the human agent is problematic from the point of view of the foundation of law and the ethics of responsibility. We therefore believe that we must keep this subject as a reference and as an irreducible basis for the use of robots. The fundamental point of our reflection is thus the following: The conduct of military operations is a serious matter. Any armed intervention must be duly considered and must at all times verify its legitimacy, legality and conformity with its purposes (which must also be legitimate ethically and legally). The use of autonomous armed robots has the peculiarity of being able to eliminate an action of its content, its purposes, and its roots and easily conceal or dilute the responsibilities associated with it. Ultimately, the use of LAWS is no longer a human action but, perversely, it tries to exonerate the initiators.

Criticism of LAWS is not based on the unworthiness of being injured or killed by a robot rather than by a human being. Indeed, a napalm bombardment by a fighter flying at low altitude is just as little respectful of human dignity as an attack with innovative LAWS. Our ethical criticism rests fundamentally on a denaturation of the action of the human person by a loss of content of human action and the refusal of the responsibilities associated with it.

We have seen conflicts created by a kind of economic logic, partly autonomous in relation to political decision-making bodies. The development of industry and military research, with their internal springs (maximization of profit, etc.), have been able to substitute themselves for genuine political will by enslaving them to their own imperatives. Today, crucial political decisions about conflicts can be automated due to the total robotization of decisions and the enslavement of human beings to machines.⁴² Gradually, one could see a loss of meaning and of the strictly anthropological content of the actions most important to the human being.

A number of specialists in the ethics and law of “algorithmic governance”⁴³ join the conclusions of our analysis, highlighting the dangers of a **lack of accountability**. The reference to a responsible human person underlies the whole system of law and ethics. As Antoinette Rouvroy points out⁴⁴:

“Many of the elements that make a person’s complexity are beyond the scope of digitization. Moreover, a decision on a person always needs to be justified by the person taking the decision, taking into account the particular situation of the individual concerned. However, automatic recommendations often work on relatively opaque logic, difficult to translate into a narrative and intelligible form. The algorithms can help the judges, but cannot dispense them from taking into account the incalculable, the non-digitizable, nor to justify their decisions with regard to this part of undecidable.

Algorithms are toxic if we use them to optimize the intolerable by abdicating our responsibilities - to hold ourselves in a position that is right in relation to our own ignorance and to make use of the collective capacities we have to make the world change. Algorithms are useful, however, when they allow us to become more intelligent, more sensitive to the world and its inhabitants, more responsible, more inventive. The choice of using them in a lazy and toxic or courageous and emancipating way is ours."

C. The human person is a being of fraternal relationship

What must be targeted at the global level is to build lasting peace among nations. This can only be done in a spirit of trust, through a genuine "ethic of fraternity".⁴⁵ This fraternity is built on human relationships, on exchanges where compassion, understanding and respect for values must take precedence over profit. A world in which autonomous machines are left to manage, rigidly or randomly, fundamental questions of the lives of human beings and nations, leads us imperceptibly to dehumanization and to a weakening of the ties that underlie the possibility of a true and lasting fraternity. What defines the human person is precisely this creative possibility to leave the frameworks (bureaucratically, legally or algorithmically) established to invent a way of restoring a place to the person who would eventually be broken by them. The risk of the robot or autonomous systems is to conceal under the performance of machines, the limits of a tool that is incapable of true creativity or real inventiveness. But this inventiveness, from the human or diplomatic point of view, is what precisely allows us to overcome the obstacles, to restore a chance to peace when all would prohibit it, or to reconcile people or nations.

A real fraternity between people regularly requires encounters and human exchanges. Ideas, rules, and laws are not enough. Dialogue and face-to-face interaction between people is essential to build confidence and to create an atmosphere of mutual understanding. Emmanuel Levinas has shown the essential character for the ethics of meeting the face of the other, which breaks the circle of the ego⁴⁶. As François Poirié rightly points out in Emmanuel Levinas' words⁴⁷: "The face of the other is what breaks the violence that is no longer seen as a murderous drive, but as carelessness, as indifference, as egoism." A conception of ethics based solely on formal systems of rules leads to a gradual confinement in a totality, to use the term of Emmanuel Levinas, which ultimately prevents the irruption of unexpected actions, unheard of, which are precisely those which mark the respect of a real otherness by opening spaces where this otherness can exist again as a brother/sister, no longer as an enemy or as a "number". The decisive point is to prevent robotization from masking increasingly the human person and the "transcendence" of his/her face. And it goes hand

Algorithms are useful, however, when they allow us to become more intelligent, more sensitive to the world and its inhabitants, more responsible, more inventive.

A world in which autonomous machines are left to manage, rigidly or randomly, fundamental questions of the lives of human beings and nations, leads us imperceptibly to dehumanization and to a weakening of the ties that underlie the possibility of a true and lasting fraternity.

“Technological and economic development that does not leave a better world and an entirely superior quality of life cannot be considered as progress”.

in hand with the idea that “technological and economic development that does not leave a better world and an entirely superior quality of life cannot be considered as progress”.⁴⁸ It is the human person, and his/her deepest reality that must remain the measure of technology and in particular of robotization.

This point is essential even when thinking about robotics in the civilian field, but it is often obliterated and forgotten. There is, in fact, a tendency in techno-science, and even in certain philosophical movements, to neglect the insistence on the unique importance of the singular encounter of the other.⁴⁹ Indeed, the depth of this encounter escapes, at least in part, a formal and positive apprehension. It is therefore crucial to think anew, the importance of the face, the body and its emotions, and the encounter, using other domains of knowledge than those of techno-sciences or empirical-formal sciences, to determine the ethical criteria allowing the regulation of a growing robotization. As Pope Francis reminds us, in the Encyclical Letter *Laudato Si'*, concerning the great challenges of mankind, those of the environment and poverty⁵⁰:

“A science which would offer solutions to the great issues would necessarily have to take into account the data generated by other fields of knowledge, including philosophy and social ethics; but this is a difficult habit to acquire today. Nor are there genuine ethical horizons to which one can appeal. Life gradually becomes a surrender to situations conditioned by technology, itself viewed as the principal key to the meaning of existence. In the concrete situation confronting us, there are a number of symptoms which point to what is wrong, such as environmental degradation, anxiety, a loss of the purpose of life and of community living.”

D. The human person is a being of compassion and forgiveness

With all the major challenges humanity is facing, an ethic based on the value of dialogue and the relationship between people who recognize themselves as brothers/sisters is more than ever needed. But the other can be or become again a brother/sister only if we can be touched deeply by what affects them, even if they have been our enemy. It is precisely in the human relationship that a deep empathy⁵¹ and a genuine compassion can be born, which can constitute a possible bulwark to the blind outburst of violence. The distancing between belligerents caused by robotic combat technologies has a well-known effect, which is to alleviate the seriousness of the damages they have caused. But this distance also has the effect of preventing the onset of an attitude of compassion or mercy⁵² that could potentially break the spiral of violence locally. It may therefore be suggested to maintain on the horizon of an evaluation of new technologies (civilian or military) this “ethic of mercy” to preserve places

where the human person can discover and be touched by the suffering of others.

Paul Ricœur's reflections on forgiveness could be summed up as follows. The human person is a being capable of forgiveness which is quite central in the dynamics of reconciliation and a return to peace. Forgiveness, which does not mean to forget, is often forgotten, but it is absolutely essential in the *post bellum* period. As Paul Ricœur says⁵³: "Forgiveness is a kind of healing of memory, the completion of its mourning; delivered from the weight of the debt, memory is released for large projects. Forgiveness gives a future to memory". Paul Ricœur also emphasizes that forgiveness escapes the law and we could add ourselves, any attempt of utilitarian formalization, because⁵⁴:

"[...] it belongs to an economy of gift, by virtue of the logic of superabundance which articulates it and which must be opposed to the logic of equivalence presiding over justice; in this respect, forgiveness is not only supra-legal but supra-ethical."

It is understood here that by breaking all logic, the inventiveness and originality of forgiveness, are not susceptible to formalization, neither algorithmic nor legal. Forgiveness implies an asymmetry and an unheard-of gratuitousness, sometimes properly "irrational", which reveals something profound and unique of the human person, that is, being able to restore faith and see as a brother/sister those who have done us wrong.

An **ethic of fraternity** is indispensable for a lasting peace. But the establishment of the conditions for the emergence of such a fraternity involves unexpected, creative, effective and symbolic gestures that can open minds and hearts and provide them with new ways of dialogue and exchange between people. To entrust the conduct of conflicts and societies totally to autonomous systems is to deprive oneself at some level or at a certain moment of this truly creative potentiality of the human person who breaks the implacable logic of violence, of utility and opens spaces for understanding. To limit the inhumanity of the war, to the extent it is possible, it is important to preserve, at the heart of the tragedies of armed conflicts, spaces where people can still be sensitive to the misery of the other being regarded as a human brother/sister and where one can risk gestures of forgiveness. This may seem paradoxical, but it is absolutely essential. If the possibilities and opportunities of being touched by the suffering of the other are eliminated, even if it is an enemy, it can lead to all the abuses and spiral into unlimited violence.

The human person is a being capable of forgiveness which is quite central in the dynamics of reconciliation and a return to peace. Forgiveness, which does not mean to forget, is often forgotten, but it is absolutely essential in the *post bellum* period.

To limit the inhumanity of the war, to the extent it is possible, it is important to preserve, at the heart of the tragedies of armed conflicts, spaces where people can still be sensitive to the misery of the other being regarded as a human brother/sister and where one can risk gestures of forgiveness.

5. Conclusion

The practical consequences of this study on the ethical evaluation of the use of robots will be outlined hereafter. If we remain faithful to an ethic based on the respect for the characteristics that are essential to the human person and which have been underlined above, we must exclude a series of robotic military technologies.

We should certainly dismiss innovative armed robots, i.e., LAWS whose autonomy is maximum, without human supervision with respect to the programming or the learning, since at some point they may run the risk of deviating from the areas of evolution or the objectives prescribed by the political or military responsible authority. It would be irresponsible to allow an autonomous weapon system to evolve, knowing that it can either learn to behave unacceptably in our eyes, or reprogram itself by redefining its own objectives, and even go beyond the limits of the scope of action envisaged. The loss or dilution of responsibility is strictly unacceptable, for it nullifies the decision of its strictly human nature. It is therefore necessary to consider prohibiting the research and development of these systems of armed robots which human agents do not pilot or supervise (a situation where the human person is totally “out of the loop”) and for which they agree not to keep the mastery of their programming and/or their learning. In the case of a system with lethal capabilities, the absence of a form of supervision is extremely problematic. When he/she accepts the use of systems capable of causing death, the human person engages his/her responsibility at a high level and cannot conceal his/her responsibilities.

In order to keep an effective accountability of the actions undertaken, (i.e. to remain the true subject of their actions) it is important to maintain control over programming and learning. The use of systems, or indeed troops, that are supervised “from the outside”, but whose reliability cannot be guaranteed, due to poor “education” or “training”, could also be seen as a form of irresponsibility.

Robots should, at all times, be able to differentiate between combatants and non-combatants without resorting to human judgment. This, however, can be problematic because recognizing the difference between a combatant and a non-combatant is sometimes very difficult. An example could be the case in which a fighter is surrendering, while still holding his gun in a non-offensive manner in his hand.

The case of robots only supervised (and not tele-operated) can pose particular ethical and legal problems. Indeed, the supervisory process leaves the machine a priority for action. But the human agent may find himself/herself in a situation where his/her time of reaction to counteract a harmful action “decided” by the machine is too short. In a number of critical but frequent situations, supervision could very well merely result in the observation by the human agent of the damage caused by the machine

he/she is supervising. The operator has virtually no time to interact with it to change its behavior. The acceptance of armed robots with supervised behaviors must therefore be taken with great caution, to the extent that supervision can be a lure. In this sense, the supervision becomes merely a *posteriori* observation of the effects produced. In these situations, supervision would be confused with cases where the human agent would be “out of the loop”, the supervision becoming, in these circumstances, practically impossible or ineffective.

One can think that there will never be a large-scale combat of robots against other robots. The belligerents will always seek to reach human beings (and more than certainly civilians) because with robots becoming less and less expensive the loss of them will have only a minor impact. That is the question from an ethical point of view. Robots are likely to be used almost invariably in asymmetric combats where a technologically advanced nation will fight groups with little or no sophisticated war material. This means that the use of LAWS or weakly supervised robots will in fact lead to a sort of confrontation between human beings on the one side, and “combatants” without faces and conscience on the other side. This situation must also be rejected because the multiplication of such “combatants” prohibits the possibility of the appearance of gestures of compassion or forgiveness. These gestures represent possible barriers, modest indeed but real, against the impulses of violence.

In the context of post-conflict operations, the presence of individuals with an attitude of genuine peace is indispensable. Trust can only be restored by showing a human and fraternal attitude. This cannot be achieved by using only robots, but there is a need of preserving a place for personal relationships. In the same way that human beings cannot live in an environment of concrete and asphalt, from which all vegetation has disappeared, similarly, human beings need to be in contact with people. In the context of a return to peace and reconciliation, the disappearance of any human mediation can prove counterproductive and generate new anxieties.

In line with our analysis, and in accordance with the ethical guidelines that we have outlined, - which fundamentally defend a coherence of all our attitudes towards what defines us as human beings - we could say that in all cases we should prohibit the development of innovative armed robots in the precise sense defined above. A “culture of peace” and an “ethic of fraternity” between people sincerely seeking to found a “community of values” can only be protected by maintaining the crucial role of a conscious and responsible human agent at the beginning of any consideration that involves the risk of suffering or death of a person.

Robots are likely to be used almost invariably in asymmetric combats where a technologically advanced nation will fight groups with little or no sophisticated war material. This means that the use of LAWS or weakly supervised robots will in fact lead to a sort of confrontation between human beings on the one side, and “combatants” without faces and conscience on the other side.

A “culture of peace” and an “ethic of fraternity” between people sincerely seeking to found a “community of values” can only be protected by maintaining the crucial role of a conscious and responsible human agent at the beginning of any consideration that involves the risk of suffering or death of a person.

Vital decisions for the human person cannot be delegated to objects. From this point of view, LAWS and any form of highly innovative armed robots are a contradiction to this requirement.

Translated into English by Rev. Fr. Antoine Abi Ghanem

NOTES

1. We refer here to the very detailed work edited by N. Bhuta, S. Beck, R. Geiss, Hin-Yan Liu, C. Kress, *Autonomous Weapons Systems. Law, Ethics, Policy*, Cambridge University Press, 2016, pp. 124-125; 327-328; 354; 364. The second part of this book is dedicated to a reflection on the meaning of the concept of autonomy; See also *Lethal Autonomous Weapons Systems. Technology, Definitions, Ethics, Law & Security* (Expert opinions delivered to the CCW Informal Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva 11-15 April 2016; R. Geiss, H. Lahmann, eds). Berlin, Federal Foreign Office. Division Conventional Arms Control (OR10); Mr. Brehm, *Defending the Boundary. Constraints and Requirements on the use of the autonomous weapons systems under International Humanitarian Law and Human Rights Law*, Academy Briefing No.9, May 2017.
2. US Department of Defense. Directive 3000.09: "Autonomy in weapon systems", 21 November 2012, 13-14 (available on the network).
3. We refer, for example, to the important article by P. Asaro, "On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making", *International Review of the Red Cross*, 94, No. 886, 2012, pp. 687-709; and to the following books: J. Galliot, *Military Robots. Mapping the Moral Landscape*, London, Routledge, 2016; *Autonomous Weapons Systems. Law, Ethics, Policy* (N. Bhuta, S. Beck, R. Geiss, H.-Y. Liu, Kress eds), Cambridge University Press, 2016; *Robots on the Battlefield. Contemporary Studies and Implications for the Future* (F. Doare, D. Danet, J.-P. Hanon, G. de Boisboissel, eds), Fort Leavenworth, Combat Studies Institute Press / Saint-Cyr Coëtquidan Schools, 2014; P. Lin, G. Beck, K. Abney, "Autonomous Military Robotics: Risk, Ethics, and Design", report to the US Department of Navy, OCT of Naval Research, 2008, California Polytechnic State University, Ethics + Emerging Sciences Group, "San Luis Obispo (available on the Web).
4. D. Dubarle, "Existe-t-il des machines à penser ?", *Revue des Questions Scientifiques*, 63ème année, t. CXXI, 1950 (5ème série, t. XI), pp. 227-228.
5. D. Dubarle, "Existe-t-il des machines à penser ?", *op. cit.*, p. 228.
6. D. Dubarle, *Ibid.* P. 230.
7. See the file: "Au-delà tu test de Turing", in *Pour la Science*, n°476, juin 2017, pp. 25-38.
8. L. Devillers, *Des robots et des hommes. Mythes, fantasmes et réalités*, Paris, Plon, 2017, p. 228, the ninth of the eleven "commands of social robotics": "You (the robot!) Will be careful that I do not confuse you with a human being."
9. J.-B. Jeangène Vilmer, "Diplomatie des armes autonomes : les débats de Genève", *Politique étrangère*, 2016/3, pp. 127 et 129.
10. See L. Devillers, *Des robots et des hommes. Mythes, fantasmes et réalités*, Paris, Plon, 2017, p. 104.
11. Ch. R. Hadlock, Six sources of collapse. A mathematical perspective on how things can fall apart in the blink of an eye, USA, MAA, 2012, pp. 15-40.
12. P. Ricoeur, "Le concept de responsabilité. Essai d'analyse sémantique" in *Le Juste*, Paris, Editions Esprit, 1995, p. 64.
13. We already encounter a complex problem of assigning responsibilities in the framework of civil technological mediations, such as banknote distributors, for example Xavier Thunis's book, *Responsabilité du banquier et automatisation des paiements*, Presses Universitaires de Namur, 1996, *Travaux de la Faculté de Droit* 19: "The interposition of a complex technical object profoundly changes the terms in which the responsibility of the banker must be posed and resolved" (p.301).

14. G. Dowek, “*Qui est responsable ? En informatique, la notion d’individu se dissout dans un continuum d’objets connectés. En cas d’accident de voiture autonome, la question de la responsabilité juridique se posera*”, *Pour la Science*, n°476, juin 2017, p. 22.
15. See, for example, the analysis of this problem in the collective work: N. Bhuta, S. Beck, R. Geiss, Hin-Yan Liu, C. Kress, *Autonomous Weapons Systems. Law, Ethics, Policy, op.cit.*, Part III: “Autonomous weapons systems and human dignity”, pp. 103-141.
16. See R. Arkin, *Governing Lethal Behavior in Autonomous Robots*, Boca Raton, CRC Press, 2009.
17. See Ch. Perelman, L. Tyteca, *La nouvelle rhétorique. Traité de l’argumentation*, t. I, t. II, Paris, P.U.F., 1958 ; Perelman objections have been presented during *Quatorzième Congrès international de Philosophie* à Vienne en septembre 1968 (*Akten des XIV Internationalen Kongress für Philosophie*, t. II, Wien, Herder, 1968, pp. 269-311). It would be useful to consult the book of R. Schmetz, *L’argumentation selon Perelman. Pour une raison au cœur de la rhétorique*, Namur, Presses Universitaires de Namur, 2000.
18. P. Navarro, L. Rodriguez, *Deontic Logic and Legal Systems*, Cambridge University Press, 2014; G. Kalinowski, *La logique des normes*, Paris, P.U.F., 1972.
19. Some authors express this very clearly. Thus, Hughes Bersini, a specialist in Artificial Intelligence, says in one of his books: “Dealing in this test of algorithmic governance, it seems to me inevitable, and this because this computational algorithm requires it, to privilege more often in turn the utilitarian version of “ethics”, since a software has to choose between several options whose effects are quantifiable (...) calculation, nothing but calculation! (Big Brother is driving you. *Brèves réflexions d’un informaticien obtus sur la société à venir*, Bruxelles, Académie Royale de Belgique, Collection l’Académie en poche, 2016, p. 17.
20. See, for example, M. Englert et al., “Logical limitations to machine ethics with consequences to lethal autonomous weapons”, 2014 (preprint: arXiv: 1411.2842v1).
21. This was used in one of the first uses of an algorithmic ethics by Michael and Susan Anderson (M. Anderson, S. Anderson, “Robot be Good. Autonomous machines will soon play a big role in our lives. It’s time they learn how to behave ethically”, *Scientific American*, October 2010, pp. 54-59) ; cfr N. McCrea, *An Introduction to Machine Learning Theory and Its Applications : a Visual Tutorial with examples*, <https://www.toptal.com/machine-learning/machine-learning-theory-an-introductory-primer> ; R. Garreta *A Gentle Guide to Machine Learning* <https://blog.monkeylearn.com/a-gentle-guide-to-machine-learning/> ; T. M. Mitchell, *Machine Learning*, New York, McGraw Hill, 1997.
22. W. Wallach and C. Allen, *Moral Machines*, Oxford University Press, 2008 ; M. Anderson, S. Anderson, “Robot be Good. Autonomous machines will soon play a big role in our lives. It’s time they learned how to behave ethically”, *Scientific American*, October 2010, pp. 54-59 *op. cit.*
23. We refer here to Aristotle: *Ethique à Nicomaque*, III, 3, and also to Paul Ricoeur: “*Le concept de responsabilité*” in *Le Juste* 1, Paris, Éd. Esprit, 1995. Tom M. Mitchell *Machine Learning* McGraw Hill. p. 2 (1997).
24. Aristotle, *Aristote, Ethique à Nicomaque*, VI, 1142 a 15-16, (translation and presentation by R. Bodéüs), GF Flammarion, 2004, p. 320. In this passage Aristotle emphasizes the fact that the judgment of “prudence” is distinguished from mathematical reasoning.
25. M. Delmas-Marty, *La justice entre le robot et le roseau in L’Homme artificiel* (sous la dir. de J.-P. Changeux), Paris, Odile Jacob, 2007, p. 246.
26. This is the case of Ron Arkin in his book: *Governing Lethal Behavior in Autonomous Robots*, Boca Raton, CRC Press, 2009.

27. "To automate these essential aspects of human judgment in the judicial process would be to dehumanize justice, and ought to be rejected in principle" (P. Asaro, "On banning ...", *op.cit.*, P701).
28. M. Delmas-Marty, *Les forces imaginantes du droit (IV). Vers une communauté de valeurs?*, Paris, Seuil, 2011, pp. 47-99.
29. See, for example, "Living Under Drones: Death, Injury and Trauma to Civilians from US Drone Practices in Pakistan", International Human Rights and Conflict Resolution Clinic de la Stanford Law School (Stanford Clinic); Global Justice Clinic of the *New York University School of Law* (NYU Clinic): <http://www.livingunderdrones.org/report/>; Amnesty International's report on drone strikes in Pakistan: Will I be next? US Drone Strike in Pakistan, October 2013.
30. See J. Galliot, *Military Robots. Mapping the Moral Landscape*, London, Routledge, 2016, pp. 165-186.
31. Even though there are already very significant studies on this subject. See, for example, "Living Under Drones: Death, Injury and Trauma to Civilians from US Drone Practices in Pakistan", International Human Rights and Conflict Resolution Clinic of the Stanford Law School (Stanford Clinic); Global Justice Clinic of the New York University School of Law (NYU Clinic) : <http://www.livingunderdrones.org/report/>; report of Amnesty International on the drones strikes in Pakistan : Will I be next ? US Drone Strike in Pakistan, October 2013.
32. See *Autonomous Weapons Systems* (N. Bhuta, ... eds), *op. cit.*, P. 139.
33. This is already true in the field of civil robotics: "The technology sector operates according to an internal logic that is not controlled by anyone. This leads to constant competition between new technologies, the uncontrolled robotization of our world and the exclusion of many people from technological advances (*Populorum communio. La communion des peuples*, Letter of the Belgian Bishops Conference for the 50th anniversary of the Encyclica "*Populorum progressio*" du pape Paul VI, Nouvelle Série n°41, 2017, p. 13).
34. See J. Galliot, *Military Robots. Mapping the Moral Landscape*, London, Routledge, 2015, pp. 65-93.
35. "One of the lessons learned from the revolutions of the 20th century is a structural change in the hearts and minds of people": JS Lucia, *Five Lessons in Christian Social Thought*, (traduit par Y. B. Vergel), 2016, p.135 (*Cinco lecciones de pensamiento social cristiano*, Madrid, Trotta, 2013).
36. See D. Lambert, "*Les robots, les Hommes et la paix*", *Revue des Questions Scientifiques*, 186 (3) (2015) 221-254 ; p. 247 for the explanation of this principle.
37. We refer here to the important book of Karol Wojtyla, *Personne et acte* (Avertissement, introduction et annotations de Aude Suramy ; traduction de Gwendoline Jarczyk), Paris, Collège des Bernardins/Parole et Silence, 2011.
38. H. Jonas, *Une éthique pour la nature* (W. Schneider, ed. ; traduction, avant-propos et notes de S. Courtine-Denamy), Paris, Desclée De Brouwer, 2000, p. 96.
39. I. Kant, *The Metaphysics of Moral* (6 :223) in *Practical Philosophy* (trans. And ed. by M.J. Gregory), Cambridge University Press, 1996.
40. P8_TA-PROV (2017) 0051 Civil Law Rules on Robotics, Texts adopted, Provisional Edition. European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law on Robotics (2015/2103 (INL)), Liability 59, f.
41. Available from the Web: 20170425_COMECE_contribution_on_robotics_COM
42. Father Dominique Dubarle o.p. had this fear very early in the article of *Le Monde* (28 décembre 1948, pp. 47-49): "*Une nouvelle science : la Cybernétique. Vers la machine à gouverner ?*".

43. See F. Musiani, "Governance by algorithms", *Internet Policy Review*, 2 (3) (2013); DOI: 10.14763 / 2013.3.188.
44. Interview of Serge Abiteboul and Christine Froidevaux with Antoinette Rouvroy on "*Big data : l'enjeu est moins la donnée personnelle que la disparition de la personne*" (<http://binaire.blog.lemonde.fr/2016/01/22/le-sujet-de-droit-au-peril-de-la-gouvernementalite-algorithmique/>).
45. Pope Francis calls for the promotion of "a genuine global ethical mobilization which, beyond differences of creed or political opinion, spreads and implements a common ideal of fraternity and solidarity, especially towards the poorest and those who are excluded" May 9, 2014, speech to the leaders of the UN).
46. As Stéphane Habib rightly says about Levinas: "The emergence or, more exactly, the gushing of the face opens up to pure responsibility-for-others" (S. Habib, *La responsabilité chez Sartre et Levinas* (Préface de C. Chalier), Paris, L'Harmattan, 1998, p. 78). L'auteur cite très judicieusement : E. Levinas, *Ethique et Infini*, Paris, Fayard 1982.
47. F. Poirié, Emmanuel Lévinas, Paris, Editions La Manufacture, 1992, p. 40.
48. M. Monet, *Emmanuel Lévinas. La relation à l'autre*, Domuni Press/ Les Presses Universitaires de l'Institut Catholique de Toulouse, 2015, pp. 81-82. Cfr E. Lévinas, *En découvrant l'existence avec Husserl et Heidegger*, Paris, Vrin, 1949 (2001).
49. Pope Francis, Encyclical Letter, *Laudato Si'* (194)
50. Ibid. (110)
51. This "empathy" is not a simple detection of external signs (by sensors of emotions for example). It requires to be able to put oneself in the place of the other to be able to imagine what it can feel. For an approach to the concept of empathy we refer to A. Berthoz, G. Jorland (ed.), *L'empathie*, Paris, Odile Jacob, 2004.
52. One is capable of an attitude of mercy who accepts to be touched in his deep being and even in his body by the misery of the other. The etymology of this term in biblical Hebrew refers to the womb of the woman, the mother who feels the pain of her child even in her maternal womb.
53. P. Ricoeur, *Le Juste*, op. cit., P. 207.
54. P. Ricoeur, Ibid., P. 206.

THE ROBOTIZATION OF THE HUMAN PERSON: FROM THE USE OF ARMED ROBOTS TO THE ADVANCEMENT OF THE “AUGMENTED SOLDIER”

A Caritas in Veritate Foundation Report by

DOMINIQUE LAMBERT

Professor at the University of Namur and Member of the Royal Academy of Belgium

Nowadays, a reflection on military robotics today must not ignore the issue of the “augmented soldier”. This issue is even more important today when looking critically at the transhumanist movement. We shall briefly describe this movement, which has been the subject of a large body of literature¹ and we will then focus on aspects of the various forms of enhancement or “augmentation” of combatants and their potential risks.

1. Transhumanism: A few characteristics

Transhumanism is the idea that the human person is only a transitory moment within evolution and that, through new developments and at will, the perceptive, cognitive and emotional capacities can be improved and, eventually, even modify the body.

This “philosophy” is based on the thesis that humankind has, by right, a total mastery of material existence, of life and death. Transhumanism is a step towards a **posthumanism**, where humankind, as we know it today, will be taken over by a new type of “augmented” being (either using machines or genetic engineering)². The possibility of such manipulation of the human species is possible by the simultaneous development of nanotechnologies, biotechnologies, information technologies and cognitive science, which are often grouped under the acronym NBIC³.

From a philosophical point of view, transhumanism is implicitly or explicitly linked to monistic, agnostic or atheistic positions⁴. It revisits a certain number of themes that can be found in religious traditions (transcendence, overcoming of death, aiming for eschatological happiness, eternal life, etc.), but it secularises them and considers them from a purely imminent point of view, i.e., without an authentic transcendence.

Transhumanism is the idea that the human person is only a transitory moment within evolution and that, through new developments and at will, the perceptive, cognitive and emotional capacities can be improved and, eventually, even modify the body.

Transhumanism is also a theory that denies the existence of human “nature” as a gift and as the basis of law or ethics. It considers the human person as a “plastic” being, capable of self-fulfilment, somehow inviting **self-induced transformation** to increase his/her knowledge, wellbeing and longevity. It should be questioned, if the reference to human nature is not recognized, what else can serve as a regulating parameter to judge the legitimacy of transformations or augmentations? The elimination of an absolute reference is of concern, as it leaves the door open to utopias, or phantoms of groups or individuals who, for partisan reasons (linked to profit, desire for power, etc.), might seek to impose augmentation on others. Eliminating this reference is therefore dangerous from a democratic point of view and it can also potentially generate a fracture within humanity, between groups that will be “augmented” and others that will appear “weakened”.

Some advocates of transhumanism draw support for their views on the impossibility of adequately separating “augmentation” from “repair” of human beings. They are perfectly right to say that certain repairs (hearing aids, dental prostheses, etc.) inevitably give rise to augmentation, because the natural constituents of the body are often less effective or less resistant than prostheses and implants. However, they obscure the fact that what makes it possible to distinguish an augmentation from a repair is precisely the reference to human nature. What makes it possible to accept the restoration of the body’s integrity, at the price of an augmentation, is precisely the compatibility with the essence of the human person. The elimination of nature, i.e., of a fundamental, immutable and objective ontological point of reference in being, can open the field to dangerous justifications exploiting a shift from the notion of repair to that of arbitrary augmentation.

Transhumanism serves as a philosophical framework underpinning research and industrial projects developed by, among others, large enterprises grouped under the acronym GAFAM (Google, Apple, Facebook, Amazon, and Microsoft). It therefore has a very significant economic component and dimension. The risk of fracture within humanity, to which we have already alluded, will potentially be doubled and accompanied by an economic divide. Effectively, the “augmented” populations will necessarily be those that will be dominant from an economic point of view. Social inequalities will probably be deepened by technologies which augment the human person.

2. The “augmented soldier”: Some examples

We can already see the intense use of robotics by soldiers today. At the same time, we can foresee a vast movement of robotization of the soldiers themselves, mainly by the

What makes it possible to accept the restoration of the body’s integrity, at the price of an augmentation, is precisely the compatibility with the essence of the human person.

augmentation of the soldier, which cannot proceed, as in the case of Lethal Autonomous Weapons Systems (LAWS), without posing deep ethical and legal questions⁵.

An “augmented” combatant is a soldier that has received the means or treatment that aim to increase performance. These means may or may not be robotic. Our goal here is not to provide an exhaustive panorama of all the ways in which the soldier of the future could be augmented. It may suffice to suggest some augmentations that are already in use or that are envisaged, to offer examples serving as an intuitive starting point for an ethical reflection on the topic.

It is certainly important to reflect on the questions raised by the augmentation of combatants, because the transhumanist movement—with the justifications it intends to provide for human transformations and the profits it expects to generate—will certainly create, in the next few years, groups that promote the introduction of a new type of soldier, changed by drugs, genetic modification or hybridization with machines. Let us now look at some examples of augmented performance and the issues they raise.

A. The change in performance may relate to endurance or physical strength as well as speed of movement

We are thinking here of the addition of robotic exoskeletons, for example. These are external devices that soldiers can wear and which would allow them to carry very heavy loads, to move more easily and for longer. These would eventually allow them to increase their strength to work on construction, demolition or to fight. We would think of a sort of hybrid fighter, in the sense of having a person coupled with a robot. This coupling could be external (by “mechanical” manipulation of the exoskeleton) or could go further by coupling and attaching cables directly between the brain and the machine (as in the case of exoskeletons used by disabled people). It is evident that this hybridization would lead to intrusions into the soldiers’ bodies, which raises crucial questions from the ethical and legal point of view and which will be revisited at a later stage.

B. Enhancement in terms of body protection

The augmentation of soldiers also implies the means to protect them, such as the bulletproof vest that is both light and capable of resisting strong impacts. This type of augmentation is not new since armour was already being used in the middle ages. What has changed is the technology that provides protection, without restricting movement or increasing the load that the individual needs to carry. We can think, for example, about the research into materials that become hard only when they are subjected to intense local pressure and that otherwise remain flexible and fluid.

An “augmented” combatant is a soldier that has received the means or treatment that aim to increase performance..

We could also think of “smart” uniforms, capable of acting as dressings in the event of injuries, or clothing for thermal or humidity regulation, for the use in extreme environments. Some of the augmentations we mentioned are for instance being used in space exploration. Space suits for instance have some of this smart clothing and external augmentations.

C. Enhancement in terms of physiological resilience or the modification of the body

Ways of “intrusively” augmenting the performance of soldiers have existed for a long time. One can think, for example, of all the possible forms of doping or drugs. During World War II, soldiers on all sides received medication that allowed them to stay awake for longer periods, face danger, combat stress or depression, feel euphoria, etc. Distribution of alcohol to combatants can also be considered as providing “doping” chemicals that enhance, up to a certain point, the capacity to “face” (even if the word is not the precise choice given that it is more a sort of anaesthetic) inhuman and stressful situations.

Here we could also think of devices that allow the control of the physiological parameters of combatants, and inject them, if necessary with the appropriate medicines (for wounds, contamination from gas, etc.). The presence of **nanorobots** capable of controlling or intervening in physiological parameters inside the bodies of soldiers could exemplify this type of augmentation. Nanorobots could be coupled to external warning systems. For example, we could imagine a nanorobot detecting the presence of harmful products like carbon monoxide coupled with an audible device for a crew member in a submarine. Evidently, we cannot ignore the fact that this type of nanorobot could be an object connected to a network and eventually, be activated by an outsider, with all the issues that this may raise. Let us think, for example, of nanorobots that, under certain circumstances, and activated by a command (either human or of automated decision algorithm), could release stimulant or euphoriant chemicals in the bodies of soldiers of an entire regiment. Worse yet, at the request of a superior, nanorobots could be remotely operated to release a lethal chemical or toxin in a special forces operative, fallen into enemy hands, who could be about to disclose critical information.

We would have here, again, some sort of **hybrid human-nanorobot organism** with external supervision of the robotics component. This, obviously, will pose the question of respect of the soldier’s freedom and autonomy to make decisions, as he would have ultimately been transformed into a “cyber puppet”.

Finally, the modifications of the combatant’s body can be envisaged from the genetic point of view. Several techniques could be used to augment the expression of genes “of interest” for certain physiological functions or to favour resistance to certain illnesses.

We would have here, again, some sort of hybrid human-nanorobot organism with external supervision of the robotics component. This, obviously, will pose the question of respect of the soldier’s freedom and autonomy to make decisions, as he would have ultimately been transformed into a “cyber puppet”.

D. Enhancement could equally be envisaged from the perceptual point of view

Combatants have already access to technologies that enhance their sensory capacities (night vision systems, etc.). Augmentations could go further by coupling a whole series of powerful sensors directly to the human brain. Today, soldiers have already access to robots or devices with a variety of sensors, but in this case the augmentation of perception could be directly linked to their own bodies. We could think of systems that emit or detect a range of waves (acoustic, electromagnetic, etc.) and that permit soldiers to expand the spectrum of their sensory information to perform both at day and night in different environments.

E. Augmentation of performance also applies to cognitive ability, decision-making, positioning and communication

We could imagine systems embedded in the brain (or heavily interfaced with the neural substrate) that could allow a direct connection with different networks to obtain information or to automatically translate certain languages.

As precursors, we could think of intelligent glasses (already existing) that project information of all kinds on the lenses. This information could also “incite” combatants by inviting them to carry out certain actions when the system has detected or inferred potential harm. This could also “inform” as well as “regulate”, since it could warn a combatant of the imminence of a forbidden behaviour.

When talking about augmentation we could also think about the insertion or injection of microchips and GPS beacons to locate and retrieve soldiers or ejected pilots from the combat zone. **Localisation** and **identification** of “ally-enemy” or “combatant-civilian” is crucial. These chips connected to body surveillance nanorobots could transfer information regarding the health of soldiers in real time and automatically initiate intervention by rescue services without them being heard.

3. Questions raised by the augmentation of soldiers

We will highlight here several ethical questions raised by combatant augmentation.

A. Loss of physical integrity

Brain implants, intrusive coupling of brain and machine or genetic modification are biological perturbations that can be irreversible. It is important to be very careful in this matter. The body is not simply an object, it is the source of expression of a person, it is the reflection of who we are. Modifying it for arbitrary, artificial or instrumental purposes

The body is not simply an object, it is the source of expression of a person, it is the reflection of who we are.

raises a crucial issue. Transhumanists often consider that the body is a sort of “plastic” reality that could be modified to their liking, or for their own projects. This idea is not self-evident, for all that could be done does not need to be done, and the acceptance of the body as a gift, and not as an object, has a profound meaning for the human person. In any event, we can assume that the augmented soldier, when they reach a certain age, will no longer want all the augmentations that were imposed on them for their missions. The possibility of restoring physical integrity is, therefore, an element to consider, even if we do not want to accept the body as a gift.

B. Loss of freedom and breach of privacy

Some combatant augmentations will result in a loss of autonomy or freedom. If nanorobots inside the soldiers’ bodies can be activated to inhibit or stimulate certain actions, this creates a major ethical problem. Effectively, they lose something important, namely the freedom of decision. Soldiers should never simply be executors. They must think about, and morally adhere to, the order received, otherwise they will be at risk of being irresponsible, like an automaton. Therefore, combatant augmentation must be regulated by the requirement to preserve the autonomy of judgment and action.

Combatant augmentation must be regulated by the requirement to preserve the autonomy of judgment and action.

Soldiers of the future will be hyper-connected combatants; therefore, all their movements, acts and gestures, will probably be picked up and recorded. The traceability of actions can be interesting from the point of view of ethics and law, given that criminal actions will be rapidly identified. Still, the recording of all words and behaviour could also be, at certain times, a violation of privacy. The intentional or accidental dissemination of opinions of a private nature could severely damage someone’s career or life, as they may find themselves forever confined to an image that does not fit in with what they are or what they have become. The life of an augmented soldier also needs to be preserved, even when their actions could be legitimately controlled in a timely manner.

C. Addiction

An old problem of augmentation is soldiers’ addiction to the doping substances that are distributed to them. This problem was encountered with troops during World War II. Taking stimulants or euphoricants for long periods can be harmful and it is important not to make future drug addicts out of soldiers. The freedom and autonomy of people need to be guaranteed. There is a risk that this will not be guaranteed if soldiers are blindly distributed drugs upon which they will become extremely dependent, even after their period of active service. We must consider that society will need to manage the lives and reinsertion of “doped” soldiers after their return to civilian life. If the doping of soldiers was to be considered normal by the armed forces, we could have serious problems

in terms of health and public security. The treatment of addictions will be an additional burden to the management of various problems associated with post-traumatic stress. This cannot be ignored, neither humanly, nor economically. The ethical reflection linked to the question of doping in high-level sport could serve to think about guidelines and avoid abuses in the military arena.

D. Exacerbated feelings of omnipotence - Overreaction and loss of a sense of boundaries

An important challenge posed by combatant augmentation is that it can give human beings an exaggerated sense of power, inciting them to use excessive or **disproportionate force**. This tendency is already apparent in soldiers that have learnt close-combat techniques and have been tempted to use their fighting skills in civilian life. Power demands to be regulated by a powerful ethic acquired through rigorous training. Still, the fascination of power could incite some people to make the wrong use of it. Moderation in the domain of the augmentation of strength is a delicate matter and it requires close attention, as it does in the case of money. For example, when someone suddenly wins a large sum of money in the lottery, they can lose their sense of moderation if they do not possess a solid inner strength. We could expect something similar with abnormal augmentation of physical strength.

Let us assume that the civilian population will not be augmented in the military sense. They will, therefore, be the most vulnerable to the eventual display of disproportionate force by augmented soldiers. One could abstractly think of fair fighting between equally augmented soldiers. In reality, however, there will probably be outbreaks of violence inflicted on the most fragile members of the population, who will pay the price for these augmentations. As in the case of urban conflicts today, the non-augmented population will probably be used as hostages or human shields by soldiers who will be better protected and augmented.

Remote operations (as in the case of drone operators or in high altitude bombing) can sometimes reduce the sense of seriousness of a situation: **“I am distant; therefore, I am not vulnerable.”** We could say that together with this, there will be a risk of losing the sense of physical limitations and control of violence: **“I am augmented; therefore, I am invincible.”** The feeling of invincibility is often dangerous, as it prevents nations from being truly aware of their “Achilles heel” and prevents them from thinking about non-bellicose means that they could use to preserve peace and security.

An important challenge posed by combatant augmentation is that it can give human beings an exaggerated sense of power, inciting them to use excessive or disproportionate force.

E. Suppression or dilution of the sense of responsibility

An augmented soldier risks having serious problems in terms of responsibility if: he/she is partly hybridized with a robot that can be controlled externally; he/she is under the influence of hyper-stimulants or hyper-euphorants; or if he/she is connected to a network of automated machines or human persons that can, at any moment, act or react to his decisions or his cerebral or bodily parameters. We may see an increase in a sort of collective behaviour of the augmented soldier. “The cloud” to which they will be connected will push them to indifferently commit acts they would never have carried out on their own. And this will be even more problematic since they may have the impression of being alone despite being conditioned by a crowd in the “cyber environment.” We need to be careful that augmentation does not transform soldiers into **unconscientious “cyber puppets”** whose cognitive and moral abilities have been anaesthetised by specific chemicals.

4. An ethical framework

The following will consider several aspects, from the point of view of ethics, regarding the issue of the augmentation of combatants.

A. Respecting the body and the ability to relate

A first point refers to the respect of bodily integrity. Any augmentation of the human person should be done in a way that is reversible and respects the human body. The idea here is that the body is not a substrate to be transformed, that it is not some mere material possession, but a medium through which the person expresses itself.⁶ Repair or restoration of the functions or parts of the body is legitimate and makes sense. A second point that is worth mentioning is the fact that making the body an instrument and a means to serve our purposes or those determined by our desires of power, which goes against its nature (*contra natura*), leads to objectifying and artificially enhancing the body in a way that could be dangerous, as it would make us prisoners of our ideas, utopias. Yet, we should stress, that these ideas and utopias could change and we could regret having gone in the direction of an irreversible transformation.

Respecting the body means not only respecting our own body, but that of others; it ultimately links to the ability to relate. Being able to establish a conversation and communicate is an essential gift to the human being, even in the worst situations of war. Dehumanization of enemies often starts by their moral or physical isolation. Avoiding eye contact with prisoners or not talking to them, denies their human status or their *alter ego*. The **transformation of combatants into war machines** could contribute to systematically removing the relational dimension that can temper violence

The idea here is that the body is not a substrate to be transformed, that it is not some mere material possession, but a medium through which the person expresses itself.

Making the body an instrument and a means to serve our purposes or those determined by our desires of power, which goes against its nature, leads to objectifying and artificially enhancing the body in a way that could be dangerous, as it would make us prisoners of our ideas, utopias.

and convey sparks of humanity in the middle of conflict. We could think here of the prohibition of fraternization between warring parties, such as the well-known episodes that occurred during World War I. This could be, not in the form of a command, but embedded or inserted in the robotic component or the augmentation of the combatants themselves.

The transformation or concealment of bodies by technological devices could contribute to create fear and/or anguish. This is sometimes sought in commando operations or in police operations, to unsettle the adversary. But the fear caused by combatants in wherein the human form has disappeared under sophisticated technological devices, could constitute a major obstacle to collecting information or establishing a dialogue necessary for a peaceful conclusion.

B. Respecting personal individuality

The individual is an ethically and judicially responsible subject. Augmentation should not dilute this individuality into a fuzzy and complex network of players; otherwise the augmented person will only be a part—with a varying degree of consciousness—of a complex machine wherein it has no control. The loss of the augmented soldier's individuality to "the cloud" into which they are connected, poses a similar question to that of the loss of the individual conscience of a person in an overexcited crowd. Augmentation should be compatible with the preservation of conscience and the person's individuality.

C. Respecting bodily limits: curbing violence and moderating its duration

In ethics, the limitations of human beings can be an asset. Their slowness can sometimes avoid hasty and dangerous decisions. Here, human weakness, the limits in terms of resistance to fatigue, could be a "natural" barrier to the duration of violent actions. The existence of chemicals that prevent sleep could incite soldiers to plan violent operations that last for too long and which would have been unthinkable without those chemicals. This illustrates that it may be important to recognize the link that could exist between "augmentation" and instigation of violence or its perpetuation. This link implies a moral responsibility for those who develop and use military augmentation. The withdrawal of the physical presence in the battlefield (as with remote combat) could entail an escalation of violence and lead to the use of unmeasured force. The combatants' fatigue, their physical resistance to stress, cold or heat, have been limiting factors that have contributed to contain or defeat warring countries. If augmented soldiers become desensitized to these factors, we could ask ourselves if conflicts might increase in intensity and last indefinitely. We could also ponder whether we would witness monstrous utilization of force against

This illustrates that it may be important to recognize the link that could exist between "augmentation" and instigation of violence or its perpetuation.

those who have not been augmented (civilian or armed forces), or who have not had the time to do so.

D. Respecting the subject: a human person must not become an object

A n essential criterion to judge the legitimacy of certain augmentations is refusing to objectify or convert human persons into tools. A human person must not be considered as a simple object or tool, without social or legal consequences. Yet, some augmentations, either of robotic or chemical nature (doping), could transform human persons into simple tools that could be remotely controlled to different degrees. We rightly rebel against the utilisation of people as kamikazes or human bombs, but some forms of augmentation that would allow the control of the combatant by their superiors, could be closely or loosely related to this kind of use. It could be argued that augmentations that could control human parameters and quickly heal soldiers could be beneficial⁷. Nevertheless, we cannot help but thinking that these systems could very quickly turn against the soldiers' health, if such devices are susceptible to being controlled from the outside by an irresponsible authority.

Respect for autonomy is one of the classic requirements of ethics, and especially of bioethics.

E. Respecting autonomy: the right to refuse augmentation

R espect for autonomy is one of the classic requirements of ethics, and especially of bioethics. In our case, it refers to respecting the autonomy of the person who could refuse augmentation that could cause health problems or moral issues. If a drug leads to the temporary suppression of consciousness or free will, the proposed recipient of that drug has the right to refuse it. Equally, if augmentation could have long-term effects on a soldier's health, they also have the right to refuse it. Surely, we will find cases where the demands of a mission will require certain augmentations or the use of specific medication (to stay awake, for example). Still, the operational interest cannot take precedence over the long-term health of soldiers. If this were not the case, we would be in the same situation as when sports organizations ignore the physical consequences for those who have had doping imposed on them. If the security of ships or airplanes and their occupants relies on a pilot that has ingested a medicine once, we can tolerate it (under certain circumstances). If the financial interests of an aviation company or the interests of a nation should rely on the slavery and long-term destruction of the pilots' or soldiers' bodies, the issue of augmentation takes a whole new twist. Refusing augmentation will not present question on the moral level, but also on a legal one, and the responsibility of the authorities could be questioned in the event of serious health problems suffered by people who have been augmented.

5. Conclusion

Even if the augmented soldier is not yet a reality, and if some of the augmentations are still considered fiction, we still need to be vigilant. A certain type of augmented soldier is already present in operating theatres and an “augmentative” mentality is taking hold—supported by a widespread transhumanist philosophy—and very active in some research centres or influential spheres. Ethicists should think in advance of the problematics caused and that is why we have reflected on this type of combatant of the future⁸.

A major ethical issue emerges here. In certain augmentation projects, the human person, or at least a large part of him/her, is transformed into an object. The robotized or augmented human person becomes progressively an instrument, a tool for combat. Eventually, the difference between a soldier and their equipment is erased, and they become equipments themselves; they also identify themselves as tools. Ethically, this obviously poses a problem, since a human being is a **subject that cannot be objectified**, under any circumstance, without suffering degradation of his/her **own dignity**. A human person, even a soldier, cannot be reduced to a simple means of combat, in the form of a hybrid: neither as a biological machine nor as a robotized life form.

In addition, this augmented human person, being an instrument of combat, is very **susceptible to manipulation** and subject to strong external constraints. Augmented soldiers risk, if we are not careful, being enslaved by powers that can—through doping or remote control—make them perform actions that they would condemn if they were not under duress.

An augmented soldier not only risks being an enslaved toy of an external authority, but a prisoner of its own desires of power, clouded by the magnitude of the new powers obtained by augmentation or robotization.

Ethicists must, therefore, reflect deeply on the risks that the augmentation of a combatant can convert them to a state of soldier-object, puppet of their superiors in the chain of command and slave to themselves. The criteria of ethical discernment lies in the refusal to objectify and enslave, and the preservation of the conscious, free and responsible individual.

To conclude, we ought to emphasise that we do not oppose all “augmentations”, but only those that make a human person an object or a slave. It is obvious that we do not criticize augmentation that could serve to better protect human beings and return quality of life. We welcome, for example, exoskeletons that allow disabled people to move; devices that help to carry heavy loads; robotic pumps that control and regulate insulin in diabetic patients; or all types of prostheses that contribute to simplify our existence or increase our quality of life. We cannot accept augmentation if it does not respect the integrity and dignity of the human being. Here we return to the **anthropological consistency principle** that we mentioned

The robotized or augmented human person becomes progressively an instrument, a tool for combat. Eventually, the difference between a soldier and their equipment is erased, and they become equipment themselves; they also identify themselves as tools. Ethically, this obviously poses a problem, since a human being is a subject that cannot be objectified, under any circumstance, without suffering degradation of their own dignity.

We ought to emphasise that we do not oppose all “augmentations”, but only those that make a human person an object or a slave.

previously regarding robots. No technological modifications carried out by human beings can, in a consistent way, be in contradiction with what humanity is at its core, with what defines it. To paraphrase a well-known expression, it would be strange if augmented soldiers become nothing more than a diminished being and suppressed in their humanity.

We could also suggest, following several authors, another way to envisage augmentation: as the **moral augmentation** of combatants, soldiers or officers. We could easily imagine that the growth of technological means both in the military and civilian domains cannot be conceived—without a risk to society—if it does not come with a meaningful growth in ethical training and questioning. The true “augmented human person” should be a human being endowed with a greater awareness of his/her responsibilities. We share the view of Brice Erbland⁹, who in the context of a very interesting article on the problematic of the augmentation of soldiers stated: “Wouldn’t the really useful augmentation be that of an ethical nature?”

It is not certain that the ethical education of combatants is keeping pace with the technological augmentation, since the progressive elimination of the human body and dignity—associated with the transhumanist and augmentative mentality—will probably lead to the loss of shared core values on which we would have expected to found relevant international laws and a minimum shared morality. It could also be because physical education of soldiers, with bodily limitations in mind, cannot be dissociated from the understanding and integration of some moral values: courage, temperance, solidarity with those who are suffering, etc. Concealment of the human body and disruption of human interactions could lead to a decline of the conscience and of moral standards. Considering this situation, it is important to be vigilant and promote a “moral augmentation” in association with technological progress.

The true “augmented human person” should be a human being endowed with a greater awareness of his/her responsibilities.

Wouldn’t the really useful augmentation be that of an ethical nature?

Translated into English by
Clemencia Licon Manzur, PhD and Rhodri P. Thomas, PhD

NOTES

1. For example: X. Dijon, *Le transhumanisme*, Namur/Paris, Fidélité, 2017, *Que penser de...?* 92; The books of G. Hottois, de l'Université Libre de Bruxelles, *Le transhumanisme est-il un humanisme?*, Bruxelles, Académie Royale de Belgique, 2014, L'Académie en poche; G., Hottois, J.-N. Missa, L. Perbal (editors) *Encyclopédie du trans/posthumanisme. L'humain et ses préfixes*, Paris, Vrin, 2015 ; J.-M. Besnier, *Demain les posthumains. Le futur a-t-il encore besoin de nous?*, Paris, Hachettes, 2009; L. Ferry, *La révolution transhumaniste. Comment la technologie et l'uberisation du monde vont bouleverser nos vies?*, Paris, Plon, 2016; M. Atlan, R.-P. Droit, *Humain. Une enquête philosophique sur ces révolutions qui changent nos vies*, Paris, Flammarion, 2012.
2. To understand this movement, we need to revisit the foundational texts: the report of 2002 Converging Technologies for Improving Human Performance. Nanotechnology, Biotechnology, Information Technology and Cognitive Science (Mihail C. Roco and William Sims Bainbridge eds.; NSF/DOC-sponsored report), Kluwer Academic Publishers (currently Springer), 2003; the European reports of 2004: A. Nordmann, Converging technologies. Shaping the Future of European Societies. A report from the High-Level Expert Group on Foresighting the New Technology Wave (http://ec.europa.eu/research/social-sciences/pdf/ntw-report-alfred-nordmann_en.pdf); Ch. Coenen, M. Schuijff, M. Smits, P. Klaasen, L. Hennen, M. Rader, G. Wolbring, European Parliament. Science and Technology Options Assessment. Human Enhancement Study. IP/A/STOA/FWC/2005-28/SC 35, 41 & 45, PE417.483, May 2009 (https://www.itas.kit.edu/downloads/etag_coua09a.pdf). We can add: "Transhumanist Declaration" (2012), The transhumanist Reader, Edited by Max More and Natasha Vita-More, edition first published 2013 © 2013 John Wiley & Sons, Inc ISBN 978-1-118-33431-7 (cited by <https://iatranshumanisme.com/a-propos/transhumanisme/la-declaration-transhumaniste/> (accessed on 16 June, 2017).
3. Also: Convergence of Knowledge, Technology and Society: Beyond Convergence of Nano-Bio-Info-Cognitive Technologies (M. C. Roco, W. S. Bainbridge, B. Tonn, G. Whitesides eds.), Berlin, Springer, 2014.
4. See: G. Hottois, *Le transhumanisme est-il un humanisme ?* (footnote 1): The large majority of transhumanists define themselves as agnostics or atheists, laics and free thinkers (...) Transhumanism is driven by an optimistic and voluntarist act of faith in the future, in human creativity and responsibility."
5. We refer here to: *Le soldat augmenté. Repousser les limites pour s'adapter* (en partenariat avec le Centre de Recherche des Ecoles de Saint-Cyr Coëquidan (CREC), DSI, Hors-Série n°45, décembre 2015-janvier 2016.
6. K. Wojtyla, The Acting Person (Translated by Andrzej Potocki and Edited by Anna-Teresa Tymieniecka), International Journal for Philosophy of Religion 13 (1):43-44 (1979).
7. See the research project concerning: "Self-healing of Body and Mind" (<http://www.darpa.mil/news-events/2015-10-05>).
8. We propose here the electronic resource: https://medium.com/law-and-policy/supersoldiers-ethical-concerns-in-human-enhancement-technologies-fa9bf1e06889?mc_cid=8ec97a3f5b&mc_eid=dc691a6a2d (ICRC Humanitarian Law and Policy, Dr Adam Henschke, "Supersoldiers". Ethical Concerns in Human Enhancement Technologies).
9. B. Erbland, "La tentation de l'Hubris", *Inflexions*, n°32, p. 54. (available electronically at: <http://inflexions.net/revue/numero-32>): "l'augmentation véritablement utile ne serait-elle pas d'ordre éthique?"

THE NEED FOR AN ETHIC OF THE RESPONSIBLE PERSON, AWARE OF HIS/HER LIMITATIONS

A Caritas in Veritate Foundation Report by

DOMINIQUE LAMBERT

Professor at the University of Namur and Member of the Royal Academy of Belgium

In his Encyclical Letter *Laudato Si'*¹, Pope Francis has highlighted problems linked to the fact that “immense technological development has not been accompanied by a development in human responsibility, values and conscience.” The robotization of war, together with the research aimed to augment the soldiers of the future, risk reinforcing these problems, to the extent that these technologies can help to dilute or obscure the responsibilities of decision-makers and restrict the freedom and consciousness of combatants under their supervision.

A machine can never be a **responsible subject**. We must then make sure to develop a great awareness of responsibility in those who are implementing robotic technologies that are becoming increasingly autonomous. The loss of responsibility or the tactical will to conceal responsibility, present a major challenge today, not only from a moral point of view, but also from a legal point of view. It is important not to reinforce the idea that human beings could be relieved of their responsibility simply because they delegated one of their decisions to a machine. It is necessary to have a **proper traceability** of the use of robotic means of combat, with a strict identification of those responsible.

A responsible person must be free and conscious. We cannot accept the “augmentation” of combatants who, through different technologies, would be transformed into war machines subject to commands. These situations can be complex. In fact, the coupling between a human person and a machine could be such that the individual risks being trapped in a network of information and influence, where its freedom to think or act is effectively suppressed. Insidious slavery linked to technological augmentation cannot be excluded, such as the ones already emerging from false statements present in social networks to which we are all connected, and which imprison and condition our thinking and spirit.

“Immense technological development has not been accompanied by a development in human responsibility, values and conscience.”

Although it is true that machines can help human agents to decide correctly in complex situations, it is also clear that keeping the person in the decision-making process can also be a guarantor of security, to avoid certain misunderstandings or misinterpretations.

Our reflections show the importance of the respect for human limitations. Although it is true that machines can help human agents to decide correctly in complex situations (for example, machines have faster reaction times and can cope with more information), it is also clear that keeping the person in the decision-making process can also be a guarantor of security, to avoid certain misunderstandings or misinterpretations. Human limitations play the role of a “fuse”, limiting destructive processes.

The limitations that we are talking about cannot be understood without deeply rediscovering what is a human person, in all his/her dimensions. Now, if intelligence is part of that definition, we can understand that it cannot be reduced entirely to an algorithm. All processes of interpretation, creativity and intuition that are part of human intelligence go well beyond the limits of algorithms. Indeed, algorithms and artificial intelligence could not be conceived or operationalized if creativity, intuition and semantic capacity were lacking. Intelligence cannot be just artificial, no matter how powerful it could be, for it is the real intelligence that is its foundation. The basis of law and ethics regarding new robotic technologies must reflect on this aspect in depth.

The human person is not only an intellect. It is also a body with its own powers, some of which come precisely from its limitations. The technologies for augmenting humankind often ignore the constraints and assets of corporality. A reflection on the augmentation of combatants should return to the profound sense of the body and its placement in the field of technology.

Intrinsic to the new combat technologies that we have mentioned throughout this reflection, we see an implicit philosophy emerge: that of a being for whom intelligence is reduced to a purely procedural and formal aspect and for whom the body no longer has real consistency. A purely mathematical intelligence deprived of a body or equipped with a transformed mechanical body—that is what the combatant of the future could be like. Is this what we want and what our democratic societies desire?

It is important to be careful, as the implicit or explicit will—associated with the research for augmenting combatants—of denying or ignoring the physical or psychological limits of the human person, could reinforce the intensity or duration of violent episodes.

A legal framework is certainly needed, when facing the challenges that the robotization of battlefields, in particular those raised by LAWS, together with those that will be generated by augmented combatants. Still, this will not suffice; we will also need an appropriate ethical reflection profoundly engrained in what constitutes the richness and *grandeur* of the human person.

This ethic could revalue the free and conscious human being, fully responsible for his/her actions, to whom corporeity reveals and affirms progressively as a precious gift, with its corresponding limitations and

assets. This revaluation of what is to be human is crucial. Indeed, today we witness a loss of confidence in humankind, as if only the machine or the new type of being (Cyborg, human-machine hybrid) was itself the saviour of humanity and its problems. We need to rediscover in human nature—seen as an asset and not as an obstacle—the resources to establish a framework of law or an ethic capable of successfully facing the challenges of today's world.

This profound attention to the human person, this respect for his/her own limitations, considered as assets, could lead to the establishment of a framework underpinning future discussions concerning the evaluation of these new military technologies, without disregarding those advances in science and technology that could guarantee health, well-being and peace.

1. An ethical screen? A common principle?

As we have seen, the use of autonomous robotics, as well as bodily or cognitive augmentation technologies, pose many ethical questions. In a multinational and multicultural context, it is difficult to get a consensus on a **reference ethical framework**. However, we think that it is possible, and desirable, to propose several questions that should be the object of reflection and on which it will be crucial to ponder, no matter the differences that could separate nations from the philosophical, cultural or political points of views. We can suggest some of the questions that have emerged from our small study. These questions could constitute the start of an “ethical check list” that allows us to highlight, and not forget, a series of crucial dangers that could be induced by military technology.

While it is clear that there could be many answers to these questions, it may be important to consider whether a commonality of mind based on the **respect of human dignity** should be sought. The respect of human dignity, even if often undefined, already plays a crucial role as a basis of national and international legal frameworks. The common principle on which a routine assessment of the uses of LAWS or of augmentation of combatants can be foreseen, could be the **principle of anthropological consistency**: Act and use technologies in such a way that these actions and these uses remain, continuously and in the long term, compatible with what we define to be a human person: with his/her body and spirit, with his/her strengths but also limitations.

A. An ethical check list in ten points

a. Predictability and reliability

Are military robots predictable and do their behaviour and fields of action remain in the zones prescribed by the responsible authorities (political or military)?

This profound attention to the human person, this respect for his/her own limitations, considered as assets, could lead to the establishment of a framework underpinning future discussions concerning the evaluation of these new military technologies, without disregarding those advances in science and technology that could guarantee health, well-being and peace.

b. Autonomy and freedom

Do human beings remain sufficiently autonomous and free, or are they progressively enslaved when they use robots or undergo augmentation?

Are we aware of the possibilities of fascination induced by machines and increase in performance linked to augmentations? Are we aware of the temptation to favour more efficient machines over human beings?

c. Individuality

Is the human person respected, or is he/she diluted in a network of machines and actors in which he/she is but one object among many? Does the human person preserve an individual capacity to consciously object and refuse to be robotized or augmented?

d. Respect of human dignity

Is the use of robots or technologies for the augmentation of human persons still at the service of human persons and coherent with their dignity (to what the human person is at the core), or is the development of robotics and augmentation beyond the definition of humankind?

e. Respect of the integrity of bodily limitations

Is the person's body respected? Are the augmentations imposed on them reversible and compatible with their limitations and bodily integrity? Does the augmented or doped person risk a long-term addiction?

f. Respect for private life

Is the person's private life respected? What is the current or potential use of data collected by autonomous robotic technology or the systems that augment combatants?

g. Effects on the civilian population

Are we aware of the effects of fear or dread induced on civilians by robotized technology or bodily augmentation of soldiers?

Are we aware of the possibility of the indirect consequences on civilians (terrorism, human shields, etc.) of a one-sided use of robotics?

h. Respect of the capacity to relate

Do robotization or augmentation maintain a place for the capacity to relate, or do they totally prohibit the expression of human feelings (compassion, forgiveness, politeness, etc.) and human relationships? Have we thought about the possible effectiveness, diplomatic or military, of human relations in the management of certain critical situations?

i. Awareness of the limitations of algorithms

Are we aware of the limitations of the algorithmic processes: bias in programming, impossibility of completely reducing semantics or prudent judgement related to an algorithmic process, etc.?

Are we aware of possible deviations from the self-learning processes of robots?

j. Awareness of responsibilities

In the use of robots, armed or not, can we clearly identify who is responsible for their actions? Can we avoid strategies for concealing responsibilities and the responsible actors made possible by the use of autonomous robots or augmented soldiers whose conduct is controlled by an authority through doping or any other technological system?

Translated into English by
Clemencia Licona Manzur, PhD and Rhodri P. Thomas, PhD

NOTES

1. Pope Francis, Encyclical Letter, *Laudato Si'* (105)

SECTION TWO

DOCUMENTS OF THE HOLY SEE ON LETHAL AUTONOMOUS WEAPONS SYSTEMS

INTRODUCTION TO THE DOCUMENTS OF THE HOLY SEE

ALICE DE LA ROCHEFOUCAULD

Director of the Caritas in Veritate Foundation

Lethal Autonomous Weapons Systems (LAWS) are a relatively new subject of discussion at the United Nations. The first mandate initiating the discussion in the framework of the Convention of Certain Conventional Weapons (CCW) was adopted in 2013.

Although this specific issue has not previously been addressed in the body of Catholic Social Teaching, its general are relevant and answer most of the questions raised by LAWS. In the coming years, one could expect more research on this very specific category of weapons, which are LAWS.

The following selected documents however represent the efforts of the Permanent Observer Mission of the Holy See to the United Nations and Other International Organizations in Geneva to urge further reflection and dialogue on such an important issue. From the beginning, the Permanent Observer Mission has been actively involved with some partner States, the International Committee of the Red Cross and the Campaign to Stop Killer Robots to propose a strong mandate to negotiate a treaty banning LAWS, as defined in the first part of this publication. This effort is ongoing.

The position of the Holy See has been further strengthened with the recent research and discussions on the creation of LAWS replacing the human person in conflict situations and placing a machine in a position to decide over life and death. From a legal and an ethical point of view this raises serious questioning: Are machines capable of replacing the human person in decisions over life and death and is this compatible with International Humanitarian Law? Can machines be responsible for the violations of international law? Ethically, can a machine replace the human capacity of moral reasoning?

For the Permanent Observer Mission of the Holy See to the United Nations and Other International Organizations in Geneva, the clear answer is a firm “no”. The human person must remain the master of robotization. New armed robots will negatively alter the nature of warfare and have a negative impact on the human family as a whole. Thus, the depersonalization and dehumanization of warfare is one of the most serious challenges posed by modern technologies.

At this stage, it should be emphasized that, contrary to many critics, the Catholic Church has always been in favor of technological progress; here we talk about technological progress that enhances good moral actions for the sake of human development but not technological progress that does bad. The Catholic Church openly encourages the development and use of technologies that respect the dignity of the human person but not those that hinder life or can be the cause for killing.

The Encyclical Letter *Laudato Si'* is the most evident and current interpretation of the position of the Holy See in praising technological progress. As stated, “we are the beneficiaries of two centuries of enormous waves of change: steam engines, railways, the telegraph, electricity, automobiles, airplanes, chemical industries, modern medicine, information technology and, more recently, the digital revolution, robotics, biotechnologies and nanotechnologies...It is right to rejoice in these advances and to be excited by the immense possibilities which they continue to open up before us, for **“science and technology are wonderful products of a God-given human creativity** (Pope John Paul II 1981)” ... Technology has remedied countless evils which used to harm and limit human beings.”¹ Thus, the Catholic Church has proven in the past centuries that it is far from being technophobic. Therefore, not all technologies are an unconditional blessing. According to His Holiness Pope Francis **“technological and economic development that does not leave a better world and an entirely superior quality of life cannot be considered as progress”**.² Technology needs to be “directed primarily to resolving people’s concrete problems, truly helping them live with more dignity and less suffering”.³

In fact, the Catholic Church is making the point that we are living, to some extent, in a culture of waste and destruction. We need to oppose this culture of self-destruction and once again place the human person at the centre of our concern. Fraternity and dialogue are the best suited means to “create community and live in peace.”⁴ More sophisticated and dehumanizing weapons will not do the job.

Technical progress in the military field is advancing at a rapid pace. It has not been accompanied, however, by an ethical and legal framework capable of successfully facing the challenges of today's world. What we need today is better people, respecting the human dignity and the development of every man and woman, and of the whole man and woman, rather than better technologies. **We are technologically strong, but morally and politically weak.** His Holiness Pope Francis states that "a technology severed from ethics will not easily be able to limit its own power"⁵. In this context, LAWS cannot be dissociated from humanitarian and ethical foundations. Hence, the dialogue between theologians, ethicists, scientists, engineers, and technologists becomes not just an interesting endeavor, but a mission to preserve the humanity and its future.

In conclusion, a "culture of peace" and an "ethic of fraternity" between people sincerely seeking to found a "community of values" can only be protected by maintaining the crucial role of a conscious and responsible human agent at the beginning of any consideration that involves the risk of suffering or death of a person. Vital decisions, which must be taken by human persons, cannot be delegated to objects. From this point of view, LAWS and any form of highly innovative armed robots are a contradiction to this requirement.

NOTES

1. Pope Francis, Encyclical Letter, *Laudato Si'* (102)
2. Ibid. (194)
3. Ibid. (112)
4. <https://www.ncronline.org/blogs/francis-chronicles/struggle-against-culture-destruction-pope-tells-teens>
5. Pope Francis, Encyclical Letter, *Laudato Si'* (136)

STATEMENT AT THE UNITED NATIONS INFORMAL MEETING OF EXPERTS ON LETHAL AUTONOMOUS WEAPONS SYSTEMS (CCW)

ARCHBISHOP SILVANO M. TOMASI, PERMANENT OBSERVER
OF THE HOLY SEE TO THE UNITED NATIONS AND OTHER
INTERNATIONAL ORGANIZATIONS IN GENEVA

13 May 2014

Mr. President,
Let me first commend you for the good preparation for this very important meeting, even if the mandate is simply to discuss in an informal setting emerging concerns around new technologies which would not only impact the way of conducting war but more importantly would question the humanity of our societies in relying on machines to make decisions about death and life.

In 2013, this Delegation expressed its deep concerns in relation with the use of drones and the troubling ethical consequences for users and victims alike.

While in many fields, autonomous technology may indeed prove beneficial to humanity, the application of autonomy to weapons technology is entirely distinct: it seeks to place a machine in the position of deciding over life and death. We are most troubled by emerging technologies of autonomous weapon systems which may move beyond surveillance or intelligence-gathering capabilities into actually engaging human targets. Good intentions could be the beginning to a slippery slope. When humanity is confronted with big and decisive challenges—from health to the environment, to war & peace—taking time to reflect, relying on the principle of precaution, and adopting a reasonable attitude of prevention are far more suitable than venturing into illusions and self-defeating endeavors.

Autonomous weapon systems, like any other weapon system, must be reviewed and pass the IHL examination. Respect for international law, for human rights law, and IHL is not optional. The Holy See supports the view that autonomous weapon systems have, like drones, a huge deficit which cannot be addressed only by respecting the rules of IHL. To comply, these systems would require human qualities that they inherently lack.

The ethical consequences of such systems if deployed and used cannot be overlooked and underestimated.

The increasing trend of dehumanization of warfare compels all nations and societies to reassess their thinking. The prospect of developing armed robots designed to engage human targets has the potential of changing the fundamental equation of war. Taking humans “out of the loop” presents significant ethical questions, primarily because of the absence of meaningful human involvement in lethal decision-making.

Mr. President,

For the Holy See the fundamental question is the following: Can machines - well-programmed with highly sophisticated algorithms to make decisions on the battlefield in compliance with IHL - truly replace humans in decisions over life and death?

The answer is no. Humans must not be taken out of the loop over decisions regarding life and death for other human beings. Meaningful human intervention over such decisions must always be present.

Decisions over life and death inherently call for human qualities, such as compassion and insight, to be present. While imperfect human beings may not perfectly apply such qualities in the heat of war, these qualities are neither replaceable nor programmable. Studies of soldiers’ experiences support that human beings are innately averse to taking life, and this aversion can show itself in moments of compassion and humanity amidst the horrors of war.

Programming an “ethical governor” or “artificial intelligence” to enable autonomous weapon systems to technically comply with the law of war in the areas of distinction and proportionality, even if possible, is not sufficient. The fundamental problem still exists: a lack of humanity, a lack of meaningful involvement by human beings in decisions over the life and death of other human beings. The human capacity for moral reasoning and ethical decision-making is more than simply a collection of algorithms. The human factor in decisions over life and death can never be replaced.

It is already extremely complex to apply the rules of distinction and proportionality in the context of war. Distinguishing combatant from civilian, or weighing military gain and human suffering, in the heat of war, is not reducible to technical matters of programming. Meaningful intervention by humans, with our unique capacity for moral reasoning, is absolutely essential in making these decisions.

Part of the justification for developing these weapons may be the idea that “if we don’t develop this technology, someone else will.” The development of complex autonomous weapon systems is likely out of the reach of smaller states or non-state actors. However, once such systems are developed by larger states, it will not be extremely difficult to copy them. History shows that developments in military technology, from crossbows to drones, give the inventing side a temporary military advantage. The inevitable

widespread proliferation of these weapon systems will fundamentally alter the nature of warfare for the whole human family.

Minimizing the risks to its own forces is understandable and legitimate. However, with no casualties or tales of horror from one side, the domestic political cost of waging war becomes less significant. This represents an important deterrent to overly-hastened military action, and is a deterrent that should not be lightly disregarded.

Autonomous weapon systems technology makes war too easy and removes its reliance on soldierly virtues. Several military experts and professional, who consider killing people a most serious matter, are deeply troubled by the idea of delegating these decisions to machines. Obviously these voices value the potential of robots to assist in bomb disposal, evacuation of the wounded, or surveying a battle scene, but the potential for robots to completely replace soldiers on the field remains of grave concern to them.

Furthermore, the delegation of the human decision-making responsibilities to an autonomous system designed to take human lives creates an accountability vacuum that makes it impossible to hold anyone sufficiently accountable for violations of international law incurred by an autonomous weapon system.

It is exactly these concerns that call for a multilateral approach to questioning the development and implementation of autonomous weapon systems. As in the case of actions like the Protocol on Blinding Laser Weapons, it is imperative to act before the technology for autonomous weapon systems progresses and proliferates, before such weapons fundamentally alter warfare into an even less humane, less human, affair.

Mr. President,

In conclusion, it is important to recognize that meaningful human involvement is absolutely essential in decisions affecting the life and death of human beings, to recognize that autonomous weapon systems can never replace the human capacity for moral reasoning, including in the context of war, to recognize that development of autonomous weapon systems will ultimately lead to widespread proliferation, and to recognize that the development of complex autonomous weapon systems which remove the human actor from lethal decision-making is short-sighted and may irreversibly alter the nature of warfare in a less humane direction, leading to consequences we cannot possibly foresee, but that will in any case increase the dehumanization of warfare.

Thank you, Mr. President.

THE USE OF LETHAL AUTONOMOUS WEAPONS SYSTEMS - ETHICAL QUESTIONS

PERMANENT OBSERVER MISSION OF THE HOLY SEE TO
THE UNITED NATIONS AND OTHER INTERNATIONAL
ORGANIZATIONS IN GENEVA

16 April 2015

1. Introduction

In today's conflicts we notice an increased presence of robotical devices. These may be aerial or sub-marine drones, reconnaissance or offensive robots, terrestrial scouting or de-mining robots, even transport robots. A robot is a system endowed with three essential components: sensors which collect data on the environment; processors to compute the obtained data, for example by means of artificial intelligence; and finally the means to carry out actions on the environment and to respond to the provoked reactions to these actions. In addition to conventional military robots, there now exist computerised robots. These systems, which are able to acquire information, process it and to intervene into the networks they are connected to, can wage genuine "cyberwars" (such as pirating important information networks, financial warfare, invasion of social networks by use of propaganda).

Today, some robots are also found in civil society due to their relatively accessible cost. This should make us cautious since, in addition to the possibility of invasions of privacy, these robots could relatively easily be armed and put to the service of terrorist organizations or criminal groups. This proliferation of small armed robots could lead to the importing and extension of situations of war or terrorism everywhere in the world, making such conflicts uncontrollable.

Whether it involves classical or computer robots, research today aims to give greater autonomy from their human agent. This means giving them the capacity to innovate in conflicts or cyber-conflicts. In other words, the research is geared towards granting them the ability to depart of their own accord from the original areas or tasks prescribed to them by the responsible authority.

Furthermore, the analysis of warfare robotics cannot be separated from an analysis of the nature of contemporary conflicts. A large part of the global population, today, lives in urban areas and this means that wars will

take place essentially in these urban contexts. It is important to be able to measure the impact and pertinence of recourse to robotic technology and the use of lethal autonomous weapon systems (LAWS) in these contexts.

This new situation means we need to carry out a new and more comprehensive analysis of warfare robotics. This could first be done by raising the issue of the legality of the use of military robots in relation to international humanitarian law (IHL). This is necessary but insufficient, for the replacement of the human decider by sophisticated and autonomous machinery raises a series of anthropological questions (What is our conception of the human being?) and ethical questions (what values are at stake in the delegation of important powers to robots, power over life and death for example?). After having briefly underlined certain points raised by the consideration of international humanitarian law, we will suggest certain landmarks enabling the guidance of an ethical questioning relative to armed robots and especially LAWS.

2. International Humanitarian Law and military robots

Is international humanitarian law compatible with the introduction of robotic weaponry endowed with a large autonomy? If we begin with the premise of *jus in bello*, the application of the principle of discrimination between combatants and non-combatants, and of the principles of proportionality, precaution and expected military advantage of an attack cannot reasonably be entrusted to robots. In fact, these tasks require the interpretation and understanding of the context and of situations which are not really programmable. In combat situations which, today, occur frequently in urban areas with a permanent presence of the civilian population, the appreciation of the legitimacy or legality of a military action is more and more complex and too subtle to be entrusted to a machine (which, for example, would be ineffective when facing moral dilemmas or questions raised by the application of the principle of double effect). It has been noted on many occasions that, following the premise of *jus ad bellum*, the risk exists that the use of autonomous weapon systems increases the probability of triggering conflicts, as they do not entail the loss of human lives on the part of the user and their production is relatively inexpensive. In addition, we cannot underestimate the risks connected to the miniaturization of surveillance robots, armed or not, in periods of tension preceding potential conflicts. Indeed, this could lead to the multiplication of incursions into an opponent's territory provoking the latter to counter-attack.

Following the premise of *jus post bellum*, we cannot underestimate the difficulties for achieving a durable peace that the use of autonomous weapons could bring after a conflict. In any case, the victor states who would use these should think of the effects produced by maintaining some

sort of robotized sword of Damocles hanging in permanence above the heads of the conquered.

In fact, miniaturized or highly autonomous robots already open the door, from the point of view of civil law, to possible important violations of privacy. But from the military point of view, they represent a permanent and generalized risk of espionage, which needs to be controlled by law. The necessity for this control is motivated by the fact that it is not possible to build a durable peace or an alliance between nations if such means of espionage, ruining mutual trust, are maintained on a permanent basis. We have recently been able to note such behaviour with episodes of cyber-espionage even between allied countries. Robotized espionage could accentuate this phenomenon.

IHL is necessary, and most agree that it is an important framework to regulate the use of armed (autonomous) robots. Nevertheless, this framework is not sufficient. Furthermore, we know that it has sometimes been possible to justify immoral conflicts by the use of “the theory of a just war”. It is therefore necessary to go beyond IHL to apprehend, in all its dimensions, the questions raised by robotic weapon systems.

3. Ethics: some landmarks

Our approach here is to draw up a list of points deserving particular attention and requiring discussion, if we want to approach the question of military robotization in an ethical way.

A. The anthropological coherence and the relational effects of robotization

When acquiring civil or military technological devices, supposed to assist in a set of difficult or dangerous tasks, human beings always have to watch out that they do not find themselves in a situation where, having entrusted a series of powers to their machines, they become slaves of their own inventions. The relevant principle here is coherence, since, if the devices supposed to free human beings become the cause of their oppression, we are faced with a profound contradiction. The momentary delegation or the permanent transfer of certain powers to machines endowed with the capacity for innovation and action have always to be measured in the light of this anthropological coherence.

The purpose of technological inventions robotized or not, is to help and free human beings.

Therefore, it is suitable to be attentive to the consequences of projects that, under the pretext of wanting to augment human capacity, to protect him from dangers or to free him from constraints, would end up in “diminishing” him and enslave him to his own tools. Otherwise, when confronted by military robots endowed with the capacity to innovate, we could end in a situation

where the technological device itself would develop strategies or tactics that the human conscience or will would not truly have ratified.

Another important ethical aspect in the use of armed robots is linked to relational effects.

Maybe, today, we are not measuring sufficiently the risks that represent the progressive replacement of human actors by machines. As the philosopher Emmanuel Levinas has shown, the face-to-face encounter with the 'other' is one of the fundamental experiences which arouses the moral conscience and responsibility. All wars lead to a regression in humanity.

But a war using solely and unilaterally technological systems, where man is absent, can increase even further this dehumanization. The presence, in the field, of human mediation can, in certain cases, open doors and offer occasions where dialogue can be resumed, where conflicts can be solved by means which are totally absent when using armed machines "without faces". This human presence permits us to be open to concrete occasions where empathy can operate, with the capacity to be able to be touched by others' suffering. On the contrary, armed machines are like psychopaths, insensitive to signs of suffering or sadness emanating from others.

Possibly, one will object that machines could be programmed to detect emotions and adopt conduct mimicking empathy. But empathy is something more than an immediate response to signs detected on the face or body of another. It is the capacity to be profoundly touched oneself by the misery of the other and to share in his burden. It has sometimes been objected that the history of contemporary conflicts has shown that under the influence of emotions or sentiments of vengeance for example, combatants can lose their cool or their reason and engage in terrible abuses, whereas programmed machines would not. However, it is demonstrable, on the contrary, that an officer or a soldier, when seriously formed in the respect of not only rules of engagement and of IHL, but also sensitive to human values, can, a fortiori, be an essential actor in reducing the dehumanization and barbarity involved in all wars.

We forget, or frequently minimize, that all weapons, from the perspective of their psychological impact, are not to be put in the same category. The Second World War has given certain flagrant examples, with the Stuka, whose siren terrorized civilian population, or the V1, whose characteristic sound, anxiously heard by the population, meant the imminent fall of the bomb. Certain weapons, by their intrinsic nature, are of a nature which provokes or augments the stress of civilian populations. This needs to be taken into account in the ethical evaluation of these weapon systems. Being flown over by planes susceptible of bombarding us is already a traumatizing experience, but the fact of being flown over in permanence by robotized machines susceptible of choosing and neutralizing targets at unexpected moments can be even more stressful. Besides, we see here the difference between the juridical and ethical aspects. A robotized weapon could meet

the principles of discrimination, proportionality and military usefulness, while producing important psychological traumas (difficult to quantify, but real) on the civilian population, and therefore raising important ethical questions.

B. The importance of prudential judgment

The removal of the human presence in the decision-making process for the selection of targets or the decision to shoot, and leaving this to an armed robot endowed with the capacity for innovation, raises fundamental questions. Indeed, to decide if this or that action is legal or legitimate on an ethical point of view, we need to refer ourselves to norms, to principles whose application to particular contexts demands evaluation and interpretation, which are not easy to translate into algorithms. The interpretative dimension operates on a double level: on one hand, the level of norms themselves (what is the sense, the spirit of the law?), on the other, the level of persons and contexts (what are the pertinent characteristic traits of the people and situations that make sense in the framework of the application of the law?).

A tradition emanating from Aristotle and St. Thomas Aquinas qualifies “prudence” as, that virtue that allows us to apply universal knowledge to particular contingent realities. It is created by experience, through the comparison of encountered situations and a capacity for interpretation which can be identified to a purely deductive or demonstrative procedure.

Classical discussions show indeed that the work of a judge cannot be reduced to the simple application of abstract norms regulated by pure formal (deontic) logic. For example, the qualification of a fact, the management of gaps or contradictions in the law (domestic or international), conflicts between norms, the clarification of terms that are too vague, or all situations where the personal interpretation of the magistrate cannot adequately be reproduced by a system of formal algorithms, without this interpretation sinking into the arbitrary or the irrational.

Furthermore, in conflicts involving interpretations of the ‘good’ or when faced with decisional dilemmas, the best programmed machines are more helpless than human beings. It is therefore important to be conscious that prudential judgment cannot be put into algorithms (for it is not merely a mechanical application of norms) and logic formalization. The experience that it presumes is in addition difficult to acquire by machines even if these are endowed with capabilities for self-learning and self-programming (by using genetic algorithms for example). Indeed, in this case, we do not see how to determine the criteria for selection of the situations the machine should remember as a pertinent basis for future behavior.

It is necessary to add here that a juridical or ethical decision leads sometimes to having to transgress the letter of universal rules to safeguard its spirit. Here too we see the limits of the use of programmed systems.

Indeed, how could we programme systems that should transgress their own rules? When the question of the robotization of armed conflicts is raised, it is important to think about the question of knowing if this “prudence”, so essential to the judgment, has not been removed or replaced by formal imitations which are not their equivalent.

C. Curb the desire for almighty power

Another important point which necessitates our attention is the fascination exercised by armed robots and the fantasies of power they mobilize. Their use can implicitly originate from the desire for power rather than from the wish to find proportional means for a justified defense. The development of “augmented soldiers”, these are combatants to whom we have conferred extraordinary capacities (perceptive, cognitive and of action) by robotized means (such as exoskeletons for example) or totally autonomous robots endowed with great capacities can originate not from military usefulness, but from dreams of power. What is human is recognized by our capacity to put a brake on our powers. Moral behavior proper to humans is situated in the position of the ‘juste milieu’ (the ‘happy medium’) which leads us to exercise force when justice and law require it, but also to renounce its use when higher values demand it. Accepting, at moments, to not exercise one’s power, to not deploy one’s “overwhelming power”, is truly human behavior. Indeed, it is that which paves the way to reconciliation, to pardon, to new solidarities, beyond all obstacles left by conflicts.¹

The nature and real greatness of man shows itself in the behaviours where the strong can make place for the weak. On the contrary, barbarity and human regression, originate in the deployment of a pure power of destruction devoid of any restrictions.

The ethical evaluation of research projects on military robotics has to take into account the differences between research motivated by a legitimate demand for security and defense, and research motivated by the sole will to realize a pure demonstration of force. This kind of demonstration would furthermore risk harming peace among nations, and generate complexes, stirring fear, rancor, jealousy and eventually, provoking in the adversary the desire to resort to similar methods, provoking a potentially uncontrollable escalation. The implementation, in a technological military system, of a pure fantasy of power, will by its very nature destabilize the balance of peace.

D. Curb the dangers of de-responsibilization

The robotized weapon, especially one which would be autonomous, raises important questions with regard to the exercise of responsibility. In the case of collateral damage, caused by weapon systems controlled by a human agent, the latter is spontaneously recognized as the one bearing responsibility. But if this collateral damage is the deed of an autonomous

machine, even if the final responsibility lies with the authority who put it into action, it will always be possible to deny it by invoking a series of malfunctions for which the authority would not be responsible (computer bugs, technical failures, design faults, scrambling of communications, etc.). It is a problem which also arises in the civil world when machines are used instead of humans. Experience shows that lawyers can use technological “screens” to construct strategies that mitigate or overshadow the identification of who is responsible.

This, in turn, could promote the use of armed robots exactly because of the impunity they permit. It would be easy to engage in combat or carry out targeted killings because the sponsors feel sheltered from direct pursuit by the law. Nevertheless, it is suitable to keep in mind the classical maxim: “*qui facit per alium, est perinde ac si facit per se ipsum*” (he who makes another do something, it is as if he had done it himself).

Just as it is intrinsically human in being able to withhold the deployment of his full power to let the other be, especially when he is vulnerable, similarly, it is intrinsically human to take responsibility for one’s actions. A doctor, like a magistrate or a military officer is not a mere technician (of the human body, the law or questions of defense). Their actions are not limited to the sole mechanical application of procedures.

The actions of jurists, as much as those of doctors or officers, are bound by rational decisions and by the taking of responsibilities which engage the whole of their person, this engagement confers to these actions an anthropological depth and density. We concur with Paul Ricoeur who characterizes the judge as some-one “humanizing the law”. In a similar fashion, we could say that, without the doctor, the judge or the officer, there is no one to humanize their particular domains of action, to bestow on these meaning and value. The delegation of important powers (economic, political, military, etc.) to machines must alert the ethical conscience, insofar as it deprives man of the power to behave authentically and where it leads to acts that we could qualify as, in the true sense of the word, “senseless”.

It might be important to underline here an important point concerning intentions. The evaluation of the responsibility of a person is concerned with his intentions. Concerning the selection of a target or the choice to shoot, we must absolutely ask ourselves the question of what were the precise intentions of the person who made the decision. This taking into account of intent is central to apply for example “the rule of double effect”, characteristic of situations where an action is simultaneously the cause of good and bad consequences. Indeed, this rule requires that one has fundamentally the intention of producing good consequences and not bad ones (even if the latter seem unavoidable). However, this notion of intent, equally important for both law and ethics, which influences greatly responsibility, cannot easily be linked to a concept or a technically

approached reality or apprehended by robotics or today's information technology.

E. Managing the effect of astonishment linked to robots' performance

Robots which would be expected to make decisions in the absence of any human supervision have programmes that bear the mark of a series of assumptions. Nevertheless, fascination for the performances of these machines can make us forget that their properties and behavior depend largely on the ideas of their makers. Let us take two examples.

To create a financial robot, a series of choices are made on how risks will be calculated. But this involves hypotheses on the evolution of the stock market, hypotheses which depend in their turn on certain presumptions. Different mathematical models will give different results and could lead to underestimating a potential yet important risk. The astonishment provoked by the achievements obtained by financial robots risks, in the long run, making us forget that their programmes cannot anticipate everything and may lead to catastrophes. These are linked, for example, to extremely rare financial events, whose occurrences are not included by the programmes in the calculation of risks.

In the military domain today, certain people envisage including, within the LAWS programmes, "ethical" programmes, thus making them into "moral machines". These are software programmes which are supposed to verify the compliance to, not only rules of engagement and principles of IHL, but also ethical norms. However, what ethical presuppositions are chosen for which software? Often, the defenders of "moral machines" use a utilitarian morality. But what justifies this choice? It is true that such a morality is founded on an optimizing calculus which can be translated into algorithms. Nevertheless, this calculation also requires measuring quantitatively the wellbeing or lack of it and, the value of this good or that person. How can we justify this? These are often ideologically driven choices. The fascination exercised by autonomous machines who would combat in our place and that would be endowed with a judgmental capacity can make us forget that the choices that underlay the creation of the software which commands these machines are contingent and that juridical or ethical judgments cannot be reduced to a mechanical application of formal rules or to calculations.

F. To maintain the control of the robotic object

One of the crucial points on which the ethicist will have to insist is the question of knowing if we can let machines whose behaviour is not entirely predictable proceed. One of the conditions to be able to use armed robots is the fact that we can be assured that it will never produce behaviour that is a priori prohibited by its user. However, this is never guaranteed, in an absolute fashion, whatever the sophistication of the software, due to

logical limitations that have been put forward by theoretical algorithmics. In addition, these complex systems are not sheltered from computer bugs and viruses and computer piracy. More than before, modern military technology is complex and susceptible of dangerous misappropriation by criminal organizations. This is reinforced by the fact that militaries depend more than before on private companies, because of the required levels of technological expertise but also because of the savings made by reducing the personnel of Western armed forces. These private companies do not benefit from the same degree of security and protection as those that operate in military bases.

The ethical evaluation of armed robots has to take into account this aspect, for it is fundamentally immoral to utilize a weapon the behaviour of which we cannot completely control. For this reason, completely autonomous robots have to be per se definitively prohibited since they will be capable of escaping from the behavioural constraints imposed by both the inventor and the user. If a weapon system has lethal capacity, it is necessary that it be maintained under absolute human supervision.

G. Concern for the economic aspects of robotization

The economic and financial stakes of the civilian and military use of robots are colossal. It is certain that the richer countries will profit more from the advantage that robotized weapons confer on them. Divisions will therefore appear between those who possess and deploy such combat technology and those who do not. Situations of injustice risk appearing as well as easy violations of the airspace of less technologically favoured countries. We will therefore reinforce already existing divisions between states.

In addition to these problems, an arms race for increasingly efficient robotized weapons systems risks leading to the degradation of resources that could be useful in other sectors such as health or education.

Ethical vigilance must also be applied to the civilian sector, as we can see that today the development of civilian robotics frequently precedes that of the military, and sometimes this leads to the civilian economy forcing the military sector to integrate such technologies even if they were not necessarily needed. The race towards profit could induce competitive industrial groups to escalate the production of increasingly sophisticated military robots making them more accessible to wider groups of actors. The relatively low cost of certain armed robots could then entice, in a near future, their rapid and uncontrollable proliferation.

4. Conclusions

This report argues that the reflections exposed herein give a few arguments showing that in the domain of the new robotized weapon systems, it is suitable to prohibit the systems which possess

lethal capacity and are at the same time capable of escaping effective control by human beings. In other words, lethal capacity cannot be bestowed upon a weapon system that could eventually, by its own initiative, leave the parameters of action that were laid down for it by the responsible authority.

If we define an autonomous weapon system as an armed system capable of innovating by adopting, by itself, behaviour that has not been foreseen and prescribed by the responsible user, it is clear that such a system can eventually escape, by definition, effective human control. It falls therefore under the scope of this prohibition.

It is necessary to note that the weapon systems targeted by this prohibition are not limited to hypothetical LAWS. Indeed, there could exist automatic or not completely autonomous weapon systems endowed with lethal capacity and eventually able to escape effective and operative control by the human agents. It is not the autonomy that is covered by the prohibition we have recommended, but rather the conjunction of potential lethality with a possible loss of effective control over these weapons systems.

It is important, to conclude, to recall, in a few words, the reasons that motivate the advocated interdiction.

A first basis for the prohibition of lethal weapons potentially escaping effective human control is the risk of a deresponsabilization or, at least, the obscuring of where true responsibility lies.

If armed robots endowed with a large autonomy provoke collateral damage, there is a risk that there will appear strategies seeking to exonerate the authorities which have employed these machines, by blaming instead technical malfunctions at different levels. The capacity to kill other persons implies an immense responsibility that can only profitably rely on a human decider and not on a technological device as sophisticated as it may be and of whom we can never truly say it is responsible.

A second basis to this prohibition is the risk of the aggravation of the dehumanization process. We know that all wars (robotized or not) are dehumanizing. However, a war that would be fought by the means of robotized machines that would not be effectively controlled by humans would be even worse. The eviction of man by the machine prevents the appearance of behaviour characterized by compassion, reconciliation, respect, which are essential to the achievement of a true peace. A true brotherhood of men and peoples cannot be acquired in conflicts where certain persons will be confronted, in an asymmetrical way, to attacks conducted by machines partially or totally subtracted from effective control by men.

Robotized wars, without human faces or people taking responsibility, risk obstructing the paths leading to a sustainable peace, which should be based on an ethic of true brotherhood among the nations.

A third basis of the prohibition is linked to the fact that the frequent delegation of important powers to machines risks depriving the political

authorities of their *raison d'être* and therefore of their capacity to act in a responsible manner. Being out-run by their machines and dazzled by their fascinating performance, these decision-makers risk not being able to decide anything, finding themselves in a paradox where the decider decides not to decide anymore. The autonomous combat machines risk, through their speed and their capacity, to dictate important military action policies.

The risks of deresponsabilization, dehumanization and depolitization induced by the use of lethal weapons removed from effective control by men are important enough so that we can envisage asking for their prohibition and the research that develops such systems.

NOTES

6. "A society founded on solely on relations of power would have nothing human: it would necessarily minimize the freedom of men instead of helping and encouraging it to develop itself and to perfect itself", St-John XIII, Encyclical Letter *Pacem in Terris*, n. 17, AAS, 55 (1963) 265.

ELEMENTS SUPPORTING THE PROHIBITION OF LETHAL AUTONOMOUS WEAPONS SYSTEMS

WORKING PAPER SUBMITTED BY THE PERMANENT OBSERVER
MISSION OF THE HOLY SEE TO THE UNITED NATIONS AND
OTHER INTERNATIONAL ORGANIZATIONS IN GENEVA

7 April 2016

An autonomous weapon system is a weapon system capable of identifying, selecting and triggering action on a target without human supervision. Among these types of systems, we can distinguish different levels of autonomy. For example, it is conceivable that weapons meet the definition that we have given, but which are not likely to re-program themselves (according to some learning, self-learning, or adaptation to a certain environment). Moreover, one can think of systems capable of learning and reprogramming (system based on genetic algorithm). This could allow an armed autonomous robot, for example, to redefine itself and the objectives that it must achieve, without the mediation of a human control.

When considering autonomous systems, it is important to specify whether one contemplates its use in controlled environments (geographical areas well identified wherein we know all the parameters: objects and persons therein and events that can occur there) or its use in open environments (i.e., not defined). It is also important to know if all the possible behaviors are unpredictable or predictable (in the case of re-programmable machines, it is certainly not the case, by definition).

If an unarmed autonomous robot works in a defined and known environment (we know exactly what this environment contains), and if all of its possible behavior is predictable (even if his particular behavior at some point and at some time, is meanwhile not predictable), it poses no particular problem (if such behavior is in accordance with the law and ethical principles that protect people and property).

In the contrary, if a robot, without human supervision, operates in an open environment, i.e., it is not controlled, and if all of its behavior is not predictable, it may be dangerous to be used even if it is not armed. Think for example of an autonomous robot that has entered a nuclear power plant

(well marked perimeter) to repair some facilities, but whose behavior is not predictable because of its ability to reprogram itself. One can easily imagine what this would mean in terms of risks. In rescue and repair operations, it is obvious that we should leave to the machine some degree of freedom, however it is important to ensure that it will not perform actions that we condemn or dread.

The definition of autonomy of an armed robot involves the following:

1) The absence of human supervision (this absence can be permanent or temporary. In the last case, there is a way to prevent the autonomous mode. A human agent can perform from outside this deactivation or an internal system to the machine can redirect the control to a human agent).

Differences between the supervised machines can appear when one specifies:

2) All the possible behaviors of the robot (behavior determined by a given algorithm or algorithms enabling learning and reprogramming).

3) The environment in which the robot operates: as specified from the point of view

- (a) Its geographical boundaries
- (b) Its contents (of persons and goods)
- and (c) All likely events.

We can characterize armed autonomous robots using these three points: (1) the degree and duration of supervision, (2) the predictability of the behavior of the robot, (3) and the characteristics of the environment in which it operates.

The “human control” of a robot with a certain degree of autonomy, in the sense that its behavior in an environment are not prescribed by a human agent, can be in fact “significant” if there is a human supervision (cf. (1) allowing in any time to disable the autonomous mode of operation), if the environment in which the robot is perfectly circumscribed and known (see (3a, b, c)), and finally, if all behaviors are well known (see (2)).

If the robot is armed with a lethal or a non-lethal weapon and that, moreover, it is not supervised or that it operates in an open environment in which events and everything in it are not fully known, it is likely to pose serious problems from the standpoint of international humanitarian law or ethics.

What are the risks of using weapons systems with a large degree of autonomy?

Besides the fact that it delegates to a machine the decision of life or death of a human being, one of the dangers is that these weapons could lead to strategies diluting or concealing true responsibilities in case of collateral damages. If armed machines provoke such damages it is easy and tempting, on the part of those who use them, to invoke technical malfunctions rather than face their responsibility. This could then lead some authorities to be less conservative, concerning precisely these collateral damages, knowing that there will be a way to protect themselves, on the pretext of failing technological mediations.

The autonomous robotic weapon necessarily induces a form of lack of accountability. It puts a distance, in a new way, between the human being and the gravity of the battle and its material and human consequences. We can already see this, for example, in the field of social networks. Violence can be unleashed in the virtual space while the actors, who are at the origin, do not realize the damage they can produce at the people who are attacked or ridiculed. The screen of technological mediations carries with it a significant risk of the occultation of our responsibilities and the concrete effect of our actions.

Another danger is the risk of hacking or the use of these systems by terrorist groups who could have recovered them. We know that complex systems are vulnerable to hacker attacks and consequently we cannot use machines with such destructive capabilities with the possible risk of misuse by malicious groups.

The development of armed autonomous robots will induce and stimulate an arms race, with its attendant costs and risks of reinforcing oppositions between nations.

The speed of development in the field of computer science and robotics is such that the balance of forces will quickly be broken. Each side will try to take advantage of rapid technical innovations in the field to gain superiority. A frenetic race may ensue. The more a military technique has the ability to advance, the higher the risk of seeing an arms race with permanent instability. This is the case of this type of technology (see the "Moore's Law" in computer science).

We should not underestimate the psychological impact of these weapons on people (already remotely operated drones pose such problems). In addition to the anxiety of the risks of nuclear, chemical or biological war, we will add the anxiety of seeing machines deciding about our life and death. As already Pope Paul VI said in his speech to the United Nations (4 October 1965): "The weapons, especially the terrible weapons that modern science has given you, even before causing victims and ruins, cause bad dreams, feeding bad feelings, create nightmares, distrust, dark resolutions; they demand enormous expenditures; they stop the projects of solidarity and useful work; they distort the psychology of peoples (No. 5)."

Far from contributing to the defense of peace, the accumulation of robotic arms could turn into a progressive incitement to war. This accumulation is all the easier as the technology of robots is relatively accessible.

One of the most common hazards associated with the use of robots is the fascination they hold for their performance. This can lead to overestimated value them and to even to the concealment of certain technological limitations. One could, in fact, believe that these machines by using their artificial intelligence capabilities, could replace the human person and meet the requirements of international humanitarian law. However, this is not the case. Why? Even assuming that the machines can correctly apply the principle of distinction between combatant and non-combatant (which does not happen by itself), it could not properly implement the principle of proportionality which requires careful judgment involving unquantifiable concepts.

The reference to the meaning of the military value of a target is also problematic. Prudence and respect for the law sometimes require going beyond the “letter” of the law and so as to interpret it according to the context to preserve the “spirit”. This ability of interpretation and going beyond the rules is not programmable. Even if a machine had reprogramming capabilities, it would still follow the rules. But humans have this ability to be able at any time to innovate out of the shackles of rigid rules, taking into account the unexpected, the unusual event. Human creativity is essential in certain circumstances, to get out of situations that threaten the human being, his dignity and the stability of societies. And it is precisely this creativity that is lacking in machines even autonomous ones. To be convinced, just think about the problem of whether it is legitimate or even possible to replace, in a court, a judge with a machine (reflecting on the fact that the mode of the exercise of justice is not reduced to the mechanical and algorithmic application of rules, even if they are programmed in a suitable logic adapted to the legal world, the deontic logic, for example).

If we want peace, we must not only avoid accumulation of arms, but we must also convert minds. “Peace must be born of mutual trust between nations, instead of being imposed on nations by the terror of weapons,” said Vatican II (*Gaudium et Spes* 81).

This confidence is based on an “ethics of brotherhood” between nations. But the accumulation of lethal autonomous weapons could undermine that trust. Indeed, can a nation really have trust in another if it has the capacity without incurring any risk, constantly to observe and hit the first with robotic arms? One could also point out that the disappearance of the human fighter will induce the disappearance of what the relationship of a person to a person and the discovery of the face of the other could provoke. A machine cannot have real empathy (this requires the experience to feel in one’s body what the other feels in his body. The machine has no real

corporeality). A machine is not open to the unexpected forgiveness and to a real possibility of reconciliation or pacification.

We must reflect on the use of technologically advanced weapons in the context of the new dimensions of the war. The battlefield has become global and extends both in the physical and virtual dimensions (cyberspace). The history of conflicts since the fifties of the twentieth century showed that this use has never resolved the conflict against determined specific groups or nations and knowing well enough the land on which they operate. It is clear that the huge spending on sophisticated weapons fails to restore peace. Quite the contrary! It appears that this sophisticated technology does not protect us against attacks and terrorism of all kinds perpetrated by people using rudimentary methods, but ready to sacrifice their lives. We should perhaps ask whether the race for technological sophistication has not reached its limits. The balance of nuclear terror has shown its limits, and with the research and development on lethal autonomous weapons, we are still in a logic that bears no fruit. We must face the facts, the race for sophisticated weaponry, including Lethal Autonomous Weapon Systems, is not able to fight against the current scourge of terrorism. The real fight is the one which will restore justice, respect for human rights, respect for minorities' rights, political participation, integral development, etc. This fight will not be won with technologically powerful weapons (robotized or not). The use of LAWS will only lead to false security and to instability. In any case, it will not establish the conditions for peace.

One could say that LAWS should not be prohibited simply because they are not able to meet the principles of international humanitarian law and the requirements of a prudent and responsible decision, but also because they lead us in a way that showed its limits. The fight against the forms of the most widespread wars now, namely terrorism, urban warfare, asymmetrical conflicts, cannot be won by increasing the firepower and sophistication of weapons. It has to happen in the field of dialogue and the common search for peaceful solutions to the real problems of our shared planet.

The refusal of LAWS must be based on the fact that even from an operational point of view, they cannot provide a lasting solution to the conflicts of today. We should draw conclusions from recent history and not be locked in a vision that is quite outdated and unrealistic. We must also be very careful about the economic logic that could push the development of these robotic technologies simply for reasons of financial gain and not for genuine needs of security and defense. For all the reasons mentioned before, we must be cautious about the research and development of LAWS. Now is the time to prevent LAWS from becoming the reality of tomorrow's warfare. The CCW should make a courageous decision of prohibiting lethal autonomous weapons like it did in the past concerning other types of weapons.

STATEMENT AT THE UNITED NATIONS INFORMAL EXPERT MEETING ON LETHAL AUTONOMOUS WEAPONS SYSTEMS (CCW)

ARCHBISHOP IVAN JURKOVIČ, PERMANENT OBSERVER
OF THE HOLY SEE TO THE UNITED NATIONS AND OTHER
INTERNATIONAL ORGANIZATIONS IN GENEVA

11 April 2016

Mr. Chairman,

The Holy See Delegation would like to express its satisfaction for the excellent preparatory work that you, Mr. Chairman, have undertaken to give every chance of success to our meeting.

The issue of lethal autonomous weapon systems is on our agenda for the third consecutive year. The Holy See had the opportunity to express its views on this important issue on previous occasions.

To respond positively to your request, Mr. Chairman, that encourages delegations to express their positions in preparation for a possible decision in the context of the CCW Review Conference in December, the Mission of the Holy See would like to briefly present a working paper in which we put forward a few arguments in favor of a collective action which seeks to prohibit the development and use of lethal autonomous weapon system.

A prevention policy seems to be the best approach. The historical experience of regulations, prohibitions or control of certain weapons shows that most often they took place after grave human tragedies. Nuclear weapons, chemical weapons, cluster munitions or antipersonnel landmines are a few good examples. The human costs are exorbitant when responses come late. Prevention is the only way to break the vicious circle of the race between technological progress of military means and their destructive force and attempts to better defend international humanitarian law.

The risks of lethal autonomous weapons systems are too numerous and important to be ignored. Besides the fact that it leaves to a machine the decision of life or death of a human being, one of the dangers is that these weapons could lead to strategies diluting or concealing true responsibilities, inducing a total lack of accountability. Instead of contributing to the defense of peace, they are turning into a progressive incitement to war.

If we want peace, we must not only avoid accumulation of arms, but we must also convert minds. Peace must be born of mutual trust between nations, instead of being imposed on nations by the terror of weapons. This confidence is based on an “ethics of brotherhood” between nations. But the accumulation of lethal autonomous weapons could undermine that trust.

We must be concerned about the use of these kinds of advanced weapons. It is clear that investing on sophisticated weapons fails to restore peace. Quite the contrary! It appears that these weapons do not protect us against attacks and terrorism of all kinds perpetrated by people using rudimentary methods, but ready to sacrifice their lives. The balance of nuclear terror has shown its limits, and with the research and development on lethal autonomous weapons, we are still in a logic that bears no fruit. We must face the facts, Lethal Autonomous Weapon Systems are not able to fight against the current scourge of terrorism and asymmetric wars. The real fight is the one which will restore justice, respect for human rights, respect for minorities’ rights, political participation and integral development. This fight will not be won with technologically powerful weapons. The use of LAWS will only lead to false security and to instability. In any case, it will not establish the conditions for peace.

For all the reasons mentioned, we must be cautious about the research and development of LAWS. Now is the time to prevent LAWS from becoming the reality of tomorrow’s warfare. The CCW should make a courageous decision of prohibiting lethal autonomous weapons like it did in the past concerning other types of weapons.

I thank you Mr. Chairman.

Imprimé en France par Romanzin SA

Dépôt légal Novembre 2017

ISBN 978-2-8399-2266-1

Lethal Autonomous Weapons Systems (LAWS) are a relatively new subject of discussion at the United Nations. In this working paper we differentiate four types of robots: controlled, supervised, autonomous and innovative robots, thus making it possible to move forward by further restricting and clarifying what the usual terminology describes as “appropriate” or “meaningful” human supervision. In particular, innovative LAWS raise a variety of perplexing legal and ethical issues, since a machine is a complex set of circuits that can never become a truly morally responsible agent. Technical progress in the military field is advancing at a rapid pace. It has not been accompanied, however, by an ethical and legal framework capable of successfully facing the challenges of today’s world. Vital decisions, which must be taken by human persons, cannot be delegated to objects. From this point of view, LAWS and any form of highly innovative armed robots are a contradiction to this requirement.

While ways of intrusively augmenting the performance of soldiers have existed for a long time, it is the elimination of an absolute reference to the human nature that is once again problematic. Our overarching ethical criticism rests fundamentally on a dehumanization and de-responsibilization of the action of the human agent. Eventually, the difference between soldiers and their equipment is erased and they become equipment themselves. Effectively, they lose their freedom of decision. We need to be careful that augmentation does not transform soldiers into unconscientious “cyber puppets”.

One of the objectives of this publication is to provoke a debate to develop an informed position in order to establish a common understanding of LAWS and augmented soldiers since they have implications and consequences on the entire human family. Therefore, the “ethical check list” outlined in this paper could be food for thought. This profound attention to the human person, this respect for his/her own limitations, considered as assets, could lead to the establishment of a framework underpinning future discussions concerning the evaluation of these new military technologies, without disregarding those advances in science and technology that could guarantee health, well-being and peace.

www.fciv.org

